



Ontologie pour la traçabilité des manipulations d'images médicales

Maria Aydee Sanchez Santana

► To cite this version:

Maria Aydee Sanchez Santana. Ontologie pour la traçabilité des manipulations d'images médicales. Web. Université de Franche-Comté, 2014. Français. NNT : 2014BESA2060 . tel-01297500

HAL Id: tel-01297500

<https://theses.hal.science/tel-01297500>

Submitted on 4 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



SPIM

Thèse de Doctorat



école doctorale **sciences pour l'ingénieur et microtechniques**
UNIVERSITÉ DE FRANCHE-COMTÉ

Ontologie pour la Traçabilité des Manipulations d'Images Médicales



María Aydeé **SÁNCHEZ SANTANA**

SPIM

Thèse de Doctorat



école doctorale **sciences pour l'ingénieur et microtechniques**
UNIVERSITÉ DE FRANCHE-COMTÉ

N° 1 0 h

THÈSE présentée par

María Aydeé **SÁNCHEZ SANTANA**

pour obtenir le

Grade de Docteur de
l'Université de Franche-Comté

Spécialité : **Informatique**

Ontologie pour la Traçabilité des Manipulations d'Images Médicales

Unité de Recherche :
Institut FEMTO-ST/DISC UMR CNRS 6174

Soutenue le 23 Octobre 2014 devant le Jury :

Jean-Christophe LAPAYRE	Directeur de thèse	Professeur à l'Université de Franche-Comté
Marie-Laure BETBEDER	Co-Directeur de thèse	Maître de Conférences à l'Université de Franche-Comté
Philippe ROOSE	Rapporteur	Maître de Conférences HDR à LIUPPA/T2I
Mohand-Said HACID	Rapporteur	Professeur à l'Université Claude Bernard Lyon 1
Christophe NICOLLE	Examineur	Professeur à l'Université de Bourgogne
Jaime CARRANZA MADRIGAL	Examineur	Professeur à l'UMSNH, Morelia, Mexique

À Cédric

REMERCIEMENTS

Je tiens ici à remercier toutes les personnes proches qui ont fait que ces travaux de thèse ont pu être menés à bien. C'est un tel plaisir d'écrire ces remerciements envers tant de personnes qui m'ont soutenue tout au long de ces trois années.

Au-delà de la formalité d'usage, c'est avec un grand plaisir que je remercie les membres de mon jury :

Monsieur Jean-Christophe Lapayre, mon directeur de thèse, et Madame Marie-Laure Betbeder, ma co-directrice de thèse, pour avoir cru en moi et m'avoir donné l'opportunité de réaliser cette thèse au sein de votre équipe malgré la distance qui nous séparait à la base. Grâce à vous, j'ai pu découvrir un nouveau domaine scientifique que sont les ontologies qui font le lien entre l'informatique et la médecine, mais également un nouveau pays et une nouvelle culture.

Monsieur Philippe Roose et Monsieur Mohand-Said Hacid qui m'ont fait l'honneur de rapporter ma thèse ainsi que pour les retours que vous m'avez fait. Et Monsieur Christophe Nicolle pour avoir accepté de faire partie de mon jury et pour l'ensemble de ses conseils techniques ainsi que pour son expertise.

Le Docteur Jaime Carranza Madrigal pour avoir bien voulu être membre de mon jury malgré le décalage horaire et l'obligation de se réveiller très tôt (ou de se coucher très tard) et pour m'avoir donné son point de vue de médecin sur le travail réalisé et fourni une validation de la partie médicale de mes travaux, en collaboration avec le Docteur Sonia Maria López Contreras et le Radiologue Fernando Sánchez Contreras.

Je tiens également à remercier chaleureusement :

L'ensemble des membres du département informatique pour leur accueil. Et plus particulièrement mes collègues et amis doctorants et docteurs : Rami et Oscar pour avoir été mes premiers amis au laboratoire et pour m'avoir offert votre amitié inconditionnelle. Merci à la TEAM des doctorants : Sébastien, Aloïs, Alban, Hadrien, Jean-Marie, Romain et les autres, pour votre amitié, votre aide avec le français et vos conseils, ainsi que pour l'ambiance quotidienne. Je les remercie pour leurs encouragements et leur bonne humeur.

Ma famille, et en particulier mes parents, à qui je souhaite adresse un immense merci pour leur soutien tout au long de ma vie, leurs enseignements et leur confiance.

Enfin, j'adresse une mention toute particulière à Cédric, mon compagnon, qui m'a donné plus que son amour. Il a su me soutenir tout au long de ces 3 années de thèse et a su me motiver chaque jour pour écrire et finir ce manuscrit à temps. Merci pour m'avoir toujours soutenue dans mes choix, pour nos discussions sans fin, pour regarder dans la même direction que moi et bien plus encore. . .

SOMMAIRE

Introduction	13
Introduction générale	13
Plan du mémoire	15
Quelques indications pour la lecture de ce rapport	17
 I État de l'art	 19
 1 Télé-applications collaboratives	 23
Introduction	23
1.1 Le travail collaboratif et les collecticiels	24
1.1.1 Modèle fonctionnel des environnements collaboratifs	26
1.1.2 Les modes de fonctionnement en collaboration	27
1.2 Le cloud computing	27
1.3 Télé-applications pour la médecine	29
1.3.1 Une croissance exponentielle depuis 2010	29
1.3.2 organisation collaborative des applications de santé	31
1.3.3 Aspect médico-légal	32
Conclusion	34
 2 État de l'art des ontologies médicales	 35
Introduction	35
2.1 Le domaine général des ontologies	36
2.1.1 Types d'ontologies	38
2.1.2 Les méthodologies de développement des ontologies	39
2.1.3 Les langages de représentation	43
2.2 Les ontologies dans le domaine médical	44
2.2.1 Taxonomies médicales	46
2.2.2 Développement des ontologies médicales	50
2.2.3 Ontologies appliquées	58

2.3	Cycle de vie d'une ontologie	61
2.4	Évaluation d'une ontologie	63
	Conclusion	65
II	Contribution	69
3	Conception des ontologies dans le domaine vasculaire	73
	Introduction	73
3.1	Les maladies vasculaires	73
3.2	Méthode de conception de nos ontologies	75
3.2.1	Répondre aux questions de compétences	76
3.2.2	Processus de développement	77
3.3	Définition de nos deux premières ontologies dans le domaine vasculaire . .	82
3.3.1	Ontologie du système vasculaire	83
3.3.1.1	Composants de l'ontologie	83
3.3.1.2	Formalismes pour la représentation des connaissances . .	84
3.3.1.3	Développement de l'ontologie	91
3.3.2	Ontologie du diagnostic	96
3.3.2.1	Représentation du diagnostic sur un modèle ontologique .	97
3.4	Définition d'une troisième ontologie dédiée à la traçabilité	102
3.4.1	Traçabilité et temporalité en télémédecine	104
3.4.2	Ontologie de traçabilité	105
	Conclusion	109
4	COOVADIS : <i>COllabOrative VAscular DiagnoSis</i>	111
	Introduction	111
4.1	Définition de la plateforme	111
4.1.1	COOVADIS : une aide à la collaboration du diagnostic	112
4.1.2	Architecture de la plateforme	115
4.2	Implémentation	118
4.2.1	Communication entre les ontologies	120
4.3	Fonctionnalités de COOVADIS	123
4.4	Exemple d'un cas clinique	125
	Conclusion	132
5	Validations en milieu clinique et premiers tests expérimentaux	133

Introduction	133
5.1 Cycle de vie des ontologies	133
5.2 Validation des ontologies	136
5.2.1 Interactions avec les experts	136
5.2.2 Types de validation	138
5.3 Modifications des ontologies	141
5.4 Premiers tests de COOVADIS	143
Conclusion	151
Conclusion et perspectives	155
Conclusion	155
Perspectives	156
Ma bibliographie personnelle	159
Bibliographie	161
Table des figures	171
Liste des tables	175

INTRODUCTION

INTRODUCTION GÉNÉRALE

CONTEXTE DE LA THÈSE

En médecine, le diagnostic est la démarche par laquelle le médecin, généraliste ou spécialiste va déterminer l'affection dont souffre le patient, et qui va permettre de proposer un traitement. Il repose sur la recherche des causes (pathologie) et des effets (symptômes) de l'affection. Un diagnostic médical efficace doit aujourd'hui intégrer des analyses multidisciplinaires tant au niveau des données que des experts : et compte tenu de la répartition géographique (par exemple de la désertification médicale) il peut être compliqué de réunir au même endroit les experts.

Depuis les années 2000, l'évolution des technologies de communication, en particulier Internet, a ouvert de nouvelles possibilités dans le domaine des applications collaboratives à distance et tout particulièrement celui du télé-diagnostic médical : par exemple un panel d'experts distants se réunit virtuellement par l'intermédiaire d'une salle d'examen virtuelle qui favorisera la collaboration afin de co-produire un diagnostic. Mais dans le domaine de la médecine, l'aspect médico-légal est crucial, et il a freiné le développement de ces pratiques à distance.

Ainsi au niveau français, ce nouveau domaine avait du mal à émerger, et ce n'est qu'en 2009, année durant laquelle les législateurs ont produit la nouvelle loi HPST (Hôpital, Patient, Santé et Territoires [HPST12]), que *l'officialisation* a permis de franchir un nouveau cap. Au sein de notre Équipe de recherche, sur la base de toutes ces évolutions est née l'idée d'une plate-forme collaborative, dédiée à la télé-médecine et mise à la disposition des professionnels de santé collaborateurs : la plate-forme collaborative se nomme CovotemTM. L'entreprise *Covalia Interactive*, qui la commercialise, a été créée en 2007. elle compte désormais 15 personnes (informatique, commercial, qualité) et hormis les différents prix (Émergence puis Innovation du Ministère de la Recherche, Innovateur ParisTech, Télé-médecine HIT Paris, . . .), l'entreprise a surtout remporté les marchés publics pour équiper les hôpitaux de Besançon, de Martinique, de la Réunion, de basse Normandie, de haute Normandie et tout récemment les hôpitaux de Paris, . . . (cette liste est très évolutive et non-exhaustive).

Covalia Interactive conçoit, développe et commercialise des solutions de téléconsultation, télé-expertise et téléassistance pour diverses spécialités médicales (axes définis dans la loi HPST, et sur lesquels nous reviendrons dans la première partie de l'état de l'art). L'entreprise a conservé un lien avec notre Équipe de Recherche en étroite collaboration avec son service R&D. Et les produits de l'entreprise doivent être en constantes évolutions : techniques pour suivre les évolutions technologiques (réseaux, terminaux, . . .), mais également en terme de réglementation (sécurisation à l'aide des cartes de professionnels de santé, compatibilité avec le DMP *Dossier Médical Personnel*, . . .). L'aspect

médico-légal et en particulier la responsabilité de l'acte est crucial et le Conseil d'État reconnaît que la recherche de responsabilité en cas de décision diagnostique partagée par plusieurs médecins complique la tâche du juge. Et dans ce cadre, Le Professeur Lucas de l'ordre des médecins [Luc12] a identifié la traçabilité des actes comme indispensable.

Au stade actuel, les différentes applications de télémédecine limitent la traçabilité à l'enregistrement des connexions des médecins identifiés par leur carte professionnelle mais les traces demeurent trop imprécises : il est impossible de savoir qui a fait quoi dans le diagnostic ? Seule la liste des intervenants est connue.

Les travaux menés dans le cadre de cette Thèse s'inscrivent dans ce contexte.

OBJECTIF DE CES TRAVAUX

Concernant les données manipulées au cours des diagnostics médicaux, pour nos recherches nous avons focalisé sur la manipulation d'images médicales en format DICOM (*Digital Imaging and Communication in Medicine*) qui est la norme mondiale pour les données d'images médicales. Il existe un modèle d'information, basé sur DICOM, pour permettre à la communauté de créer des bases d'archivage médicales qui soient interopérables : par exemple les serveurs PACS (*Picture Archiving and Communication System*).

Ces serveurs ont des capacités de stockage très conséquentes : on trouve couramment dans les hôpitaux des serveurs de plus de 20 Terabytes. Pour un seul patient, le nombre et la taille des données sont déjà très importants. Par expérience lors de l'utilisation du logiciel CovotemTM au sein des hôpitaux, nous avons pu observer qu'un CT SCAN Cérébral peut représenter 570 Mo, ou bien qu'une IRM peut représenter 100 Mo. Des dossiers médicaux peuvent atteindre, s'ils combinent plusieurs données d'imagerie, une taille de l'ordre du Gigaoctet : 1.2 Go pour un dossier traité en RCP (réunions de concertation pluridisciplinaire), 3 Mo pour la dermatologie, 100 Mo pour la neurologie, plus de 500 Mo en cardiologie, ...

La manipulation des dossiers devient alors très rapidement très coûteuse, et donc plus particulièrement la traçabilité des manipulations est très lourde. L'objectif de cette Thèse est donc de trouver des moyens de l'optimiser. Et nous avons ainsi orienté nos recherches vers le domaine des ontologies.

Les ontologies sont devenues très populaires dans plusieurs domaines de recherche : les ontologies encodent un domaine de connaissances en une description formelle. Cette connaissance offre différentes possibilités de raisonnements, lexiques et sémantiques qui ne sont pas offerts par les bases des données. Les ontologies sont employées dans l'intelligence artificielle, le Web sémantique, la génie logiciel, l'informatique biomédicale ou encore l'architecture de l'information comme une forme de représentation de la connaissance d'un monde ou d'une certaine partie de ce monde, ...

Dans ce travail, nous avons souhaité utiliser les ontologies, non seulement afin d'optimiser la recherche d'informations pertinentes dans les diagnostics, mais également pour ajouter des paramètres d'ontologie supplémentaires (temporels) permettant la traçabilité. Nous proposons de développer des ontologies dédiées à DICOM qui permettront d'unifier et de rendre explicite l'ensemble des entités et des relations clés dans DICOM, dans un format à la fois utilisable pour l'homme et interprétable par la machine. La création d'un

système de traçabilité des données du dossier (Patient Maladie cardiovasculaire) prévoit donc l'intégration de multiples éléments liés du patient (antécédents médicaux), l'imagerie médicale (scanner, radiologie conventionnelle, résonance magnétique, doppler, ...) et un nouveau processus d'identification directe des diagnostics sur DICOM.

Les applications de télémédecine sont présentes dans de très nombreux domaines de la médecine : comme la dermatologie, la radiologie, la neurologie, la cardiologie Dans le cadre de ces premiers travaux qui utilisent les ontologies pour la traçabilité, nous ne pouvions pas traiter la médecine dans son ensemble (toutes disciplines confondues) : de part les collaborations et les contacts dans le milieu médical entretenus par notre Équipe de Recherche, il a donc été décidé de cibler particulièrement le domaine cardiovasculaire.

Pour le développement, l'évaluation et la validation de la partie médicale, les travaux ont été réalisés en collaboration avec trois personnalités de Morelia au Mexique ainsi qu'avec un élève ingénieur français :

- le Professeur Jaime Carranza Madrigal de l'*Université Michoacana de San Nicolás de Hidalgo*, et cardiologue à l'Hôpital Civil,
- M. Fernando Sánchez Contreras médecin au *Centro Union*,
- Mme Sonia López Contreras, Directrice de la clinique UPAC (*Unidad de Prevención y Atención Cardiométabólica*),
- et Jean-Denys Maréchal élève ingénieur de troisième année à l'école ISIFC (*Institut Supérieur d'Ingénieurs de Franche-Comté*) en Génie Biomédical.

Ainsi nos ontologies ont donc été conçues en deux langues : ontologies en langue espagnol pour l'évaluation des cardiologues mexicains de Morelia, et ontologies en langue anglais afin d'être également utilisables en Europe.

PLAN DU MÉMOIRE

Comme le montre la présentation du contexte (Fig. 1), cette Thèse s'inscrit à la frontière de trois domaines : deux informatiques dont relèvent ces travaux de recherche (le travail collaboratif^❶ et les ontologies^❷) et un médical ^❶ qui est le fond applicatif des travaux.

Dans la première partie de ce mémoire, nous présentons les états de l'art qui ont été nécessaires à la mise en place de ce travail : travail collaboratif (télé-applications) et ontologies.

Dans un premier chapitre, nous mettons en avant les évolutions qui ont eu lieu ces dernières années du point de vue des télé-applications collaboratives, de leur architecture, en focalisant plus particulièrement sur les télé-applications dans le domaine de la médecine.

Le seconde chapitre est consacré à l'état de l'art des ontologies, tout d'abord d'un point de vue général (méthodes de conception, outils et langages) puis plus spécifiquement dans le domaine médical. Pour terminer, les dernières sections de ce chapitre exposent deux points essentiels qui sont le cycle de vie des ontologies et leur évaluation.

La deuxième partie de ce mémoire constitue le cœur de notre contribution. Nous y présentons notre plateforme COOVADIS (*COllabOrative VAscular DlagnoSis*) qui s'appuie sur trois ontologies originales que nous avons élaborées, et qui ont été ensuite

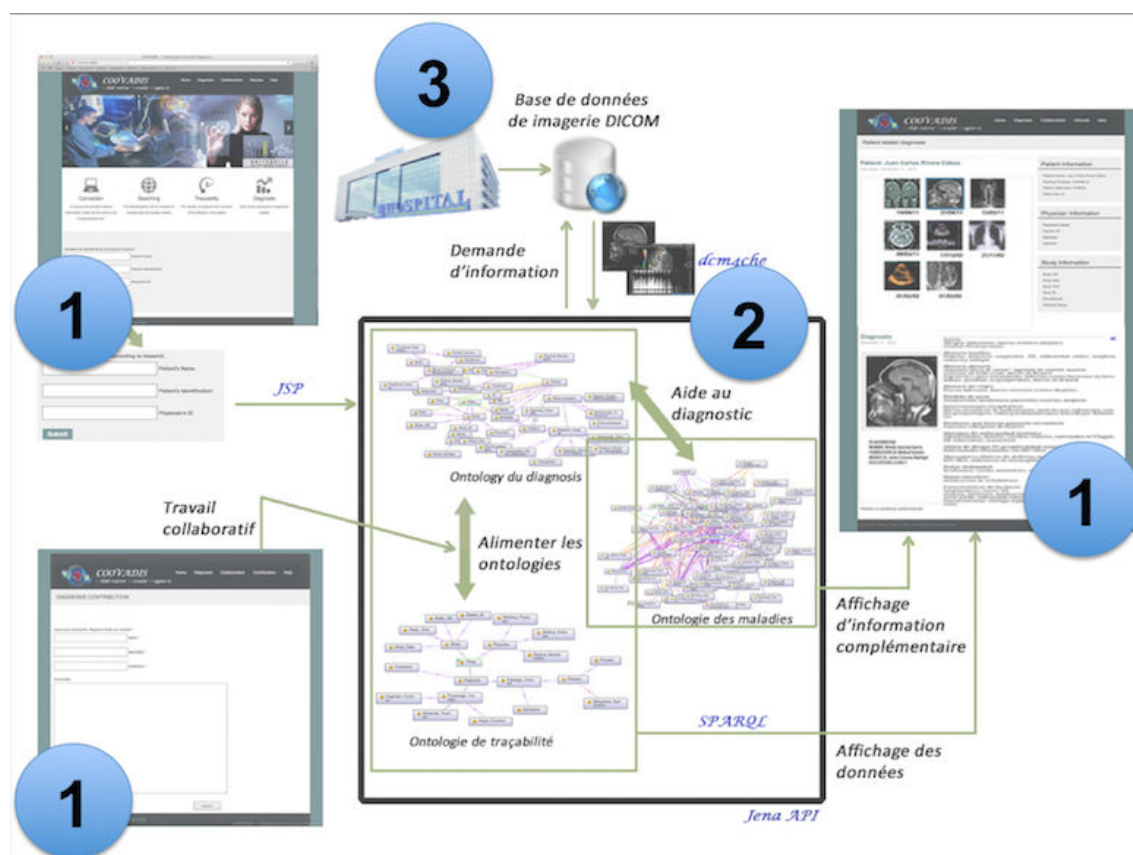


FIGURE 1 – Schéma global d'organisation des travaux de recherche

implémentées en mode SaaS (*Software as a Service*) sous la forme d'un serveur WEB, et validées d'un point de vue théorique et avec un test en milieu clinique.

Dans le troisième chapitre, nous définissons les trois ontologies essentielles pour développer la traçabilité des diagnostics dans le domaine cardiovasculaire : tout d'abord les deux ontologies liées au domaine vasculaire (ontologie du système vasculaire, et ontologie du diagnostic), puis l'ontologie de traçabilité.

Puis nous définissons dans le chapitre 4 notre nouvelle plateforme *COOVADIS COllaborative VAScular DiagnoSis* créée autour de ces trois ontologies. Développée en mode SaaS *Software as a Service*, elle permet de suivre et de tracer les diagnostics élaborés par un staff de médecins sur un patient atteint d'une pathologie cardiovasculaire, offrant ainsi la traçabilité des actes.

Le dernier chapitre de cette partie expose les aspects validation théorique, ainsi qu'une validation par un essai en milieu clinique. Nous présentons également les premiers résultats expérimentaux obtenus à l'aide de notre plateforme.

Nous pouvons préciser que dans cette partie nous présentons un cas réel « Mr J. » sur lequel des diagnostics ont été réalisés (deux Accidents Vasculaires Cérébraux successifs, et le suivi du patient). Ce patient anonyme est notre *fil rouge* et permet de montrer sur un cas concret et complet l'intérêt de notre plateforme dans le suivi et la traçabilité de l'acte médical.

En conclusion de ce mémoire, nous exposons les différents apports de nos contributions ainsi que les évolutions possibles qui pourront faire l'objet de futurs thèmes de recherche.

QUELQUES INDICATIONS POUR LA LECTURE DE CE RAPPORT

Le plan est organisé sur quatre niveaux : Parties, Chapitres, Sections et Sous-Sections. Il est utile de savoir que la Partie II, présentant les contributions et résultats, est indépendante de la première Partie.

Les résultats de nos travaux ont pour la plupart été publiés. La liste des publications personnelles est donnée avant la bibliographie placée en fin du document.

I

ÉTAT DE L'ART

INTRODUCTION

Ce sujet de Thèse se situe à la croisée de plusieurs domaines de recherche dont les deux plus importants sont les applications distribuées pour le télé-diagnostic et les ontologies.

Avec le développement fulgurant des réseaux (réseaux Wifi, 3G, 4G, . . .), la puissance des nouveaux processeurs, et la banalisation des périphériques audio et vidéo (tout ordinateur portable possède maintenant « *obligatoirement* » une webcam, un micro et un haut-parleur), la production d'outils de communication pour favoriser les échanges à distance a explosé.

Les outils grand public comme Skype, Oovoo par exemple, permettent des échanges audio et vidéo, mais ils demeurent incomplets et peu fiables : pas de garantie de qualité de service, peu ou pas de *partage* de données (ou partage de données mais sans garantie de cohérence par exemple), pas de tolérance aux pannes, pas d'adaptabilité, pas de traçabilité, . . .

Depuis les années 2000 environ, différentes communautés de recherches se sont fédérées autour du CSCW (*Computer Supported Collaborative Work*). Ce domaine, malgré tout encore très récent, regroupe des chercheurs issus de l'informatique distribuée, des sciences humaines, de la sociologie de l'ethnologie, de l'ergonomie, . . .

Dans ce domaine très large, on peut distinguer les télé-applications collaboratives qui sont des outils de travail à distance pour lesquels il faut offrir des services et des outils garantissant la cohérence du travail à distance, l'*awareness* (outils permettant la prise de conscience des autres pour favoriser les interactions distantes), des interfaces distribuées, des outils d'adaptabilité, . . .

L'équipe dans laquelle les travaux de cette thèse ont été menés étudie plus particulièrement l'algorithmique distribuée pour les télé-applications de diagnostic médical à distance : appelée télé-diagnostic médical. Et l'un des nouveaux challenges est la traçabilité. En effet, l'aspect médico-légal est très important et notamment lors d'erreur de diagnostic, il est donc nécessaire de pouvoir tracer, de manière fiable et facilement exploitable, le travail de chaque acteur du travail collaboratif de télé-diagnostic.

Les masses de données dans ce domaine sont très importantes : la taille d'un dossier peut représenter de quelques méga octets (tomographie pour la neurologie par exemple) à quelques Gigaoctets (pour une mammographie par exemple une coupe représente 50 Mo) voire même de l'ordre du Téraoctet (en microscopie, une série de coupes peut représenter 1 To). Lorsque ces masses de données sont manipulées, partagées, . . . il devient rapidement très lourd et coûteux de tracer le diagnostic. Il nous est apparu que l'utilisation des ontologies nous permettrait un stockage et une restitution d'information performants et pertinents.

Comme nous le verrons dans la suite de ce document une ontologie, dans le domaine de l'ingénierie des connaissances, est une spécification formelle, explicite d'une conceptualisation partagée [Stu98]. Dans le domaine médical, l'ontologie est à la

frontière des sciences de l'information et des sciences médicales, et elle vise à unifier la compréhension des mécanismes d'interprétation et de raisonnement médical. Par exemple, grâce à l'utilisation de l'ontologie, lors d'un diagnostic médical le langage et les concepts utilisés pour écrire le compte rendu de l'acte ne varient pas (ou peu de manière acceptable) d'un praticien à un autre. Dans notre cas, la conceptualisation partagée est l'acte médical, et si le langage et les concepts utilisés sont suffisamment formels et explicites, la recherche d'information, par exemple pour la traçabilité, sera performante et pertinente.

L'état de l'art qui suit est donc logiquement articulé en deux sections. La première présente l'état de l'art des télé-applications collaboratives et en particulier des applications de télé-diagnostic médical. Ces applications distribuées de nouvelle génération nécessitent de respecter certaines contraintes en terme de sécurité, de cohérence des données, de sûreté Notons que cette Thèse ne propose pas de contribution dans le domaine de la collaboration, mais les ontologies, qui seront proposées, seront intégrées à un serveur de collaboration.

Dans le domaine médical en particulier, mais pas uniquement dans ce domaine, il est nécessaire de pouvoir tracer les actions effectuées par les différents acteurs dans le système collaboratif, et c'est dans ce contexte émergeant que s'inscrivent les travaux de cette Thèse. Pour tracer il est nécessaire de bien définir la structure des informations de traçage, et ainsi la deuxième section étudie l'état de l'art des ontologies plus particulièrement utilisées dans le domaine médical. Notre démarche a été descendante : nous sommes partis de l'état de l'art assez large des ontologies pour ensuite donner les caractéristiques et les spécifications des ontologies dédiées au monde médical. Pour finir cette section nous étudions la notion de temps dans les ontologies : pour tracer un diagnostic collaboratif, il faut structurer l'information mais également la dater pour ordonner les actions de diagnostic. En effet, pour des pathologies « dites lourdes » (de type cardiovasculaire par exemple), différents examens sont nécessaires et même pour un seul examen plusieurs médecins peuvent intervenir à quelques heures, minutes ou secondes près, et il est important d'ordonner ces interventions afin de suivre la démarche diagnostique à des fins médico-légales ou d'enseignement par exemple. Ainsi, nous présentons les ontologies et les modélisations, issues de la littérature, qui permettent de tenir compte de la temporalité.

TÉLÉ-APPLICATIONS COLLABORATIVES

Les travaux relatifs à cette Thèse ne sont pas directement rattachés, en terme de contribution scientifique, au travail collaboratif. Ce domaine est le cadre global puisqu'il s'agit d'utiliser les ontologies afin de faciliter la traçabilité dans les téléapplications collaboratives.

Il était donc important, avant d'aborder l'état de l'art principal de ce document sur les ontologies, d'étudier au préalable les applications collaboratives.

INTRODUCTION

Durant ces dernières années, l'accroissement des performances informatiques au travers d'Internet a favorisé la collaboration de personnes travaillant dans le monde entier : optimisation des bandes passantes, performances accrues des processeurs, optimisation des codecs, accroissement de la taille des mémoires... Ainsi, de nouvelles technologies se sont appuyées sur ces capacités afin de développer des outils et applications collaboratives qui permettent à plusieurs utilisateurs de travailler simultanément sur des ressources communes (médias, documents ...). Le principe de la plupart de ces applications est le partage des données, audio et/ou vidéo au sein d'un groupe de personnes.

Ce nouveau domaine dont l'essor réel s'est réalisé dans les années 2000, est connu sous le nom de CSCW (Computer Supported Collaborative Work), et il regroupe des compétences non seulement dans les sciences informatiques mais également dans les sciences humaines comme l'ergonomie, la psychologie, la sociologie, l'ethnologie, ...

Le but des chercheurs en applications collaboratives est de permettre la communication à distance, afin d'éviter l'éventuelle nécessité de déplacements physiques pour travailler ensemble. Le développement de la télé-informatique a permis l'apparition de nombreux logiciels pour le travail de groupe : les domaines d'utilisation les plus courants sont la télé-maintenance, la télé-médecine, et plus généralement le domaine des télé-applications. Les précédents travaux de recherche réalisés au sein du laboratoire d'informatique de Franche-Comté (Département d'informatique des systèmes complexes, Femto-ST/DISC) [Aup10, Aup12] ont permis l'amélioration des techniques de partage de données (accès concurrents, cohérence des données, protocoles de gestion de la concurrence) ainsi que l'utilisation de l'adaptabilité des médias dans ces applications. Ces développements ont permis de rendre plus fiables et performantes les nouvelles applications (audio/vidéo conférences, télé-réunions, partage de données et images, ...). L'application de ces résultats de recherche est réalisée au sein de l'entreprise Covalia

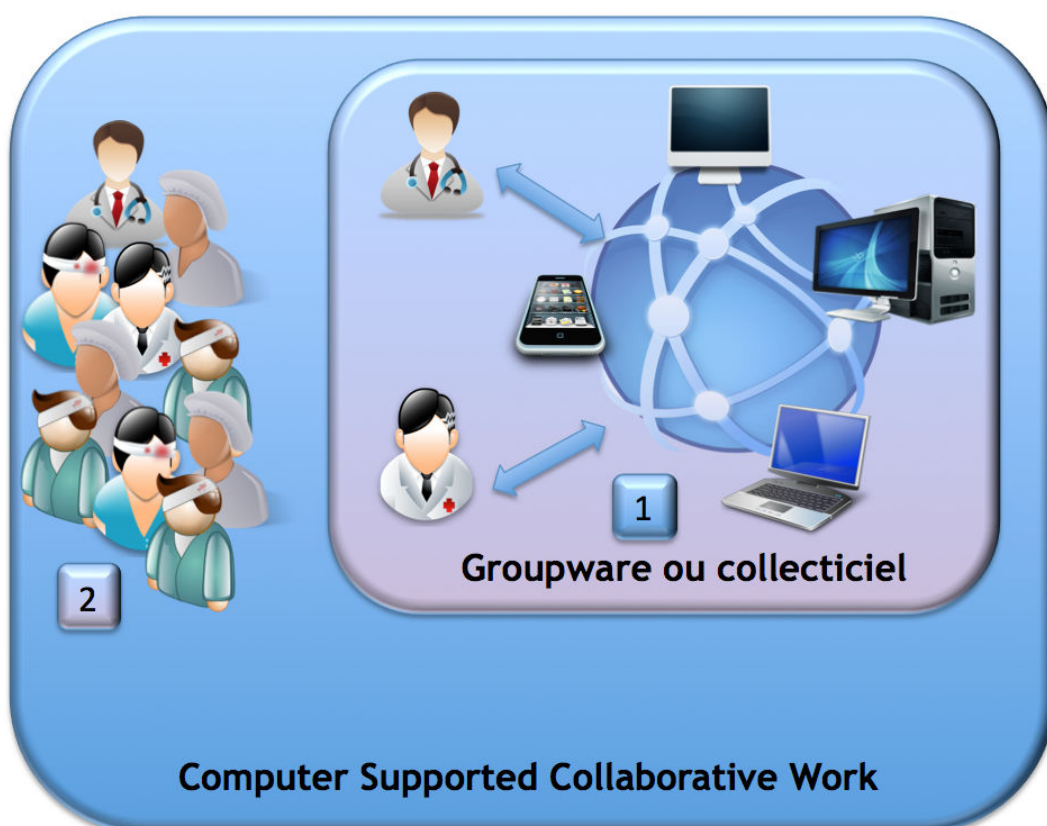


FIGURE 1.1 – Place du collecticiel dans le domaine du CSCW

interactive¹ (startup issue de l'équipe de recherche), et ainsi de nouveaux modules du logiciel CovotemTM intègrent ces résultats.

Dans ce chapitre, nous dressons l'état de l'art des applications de travail collaboratif, aussi appelées collecticiels, puis nous développons plus particulièrement nos recherches sur le domaine des applications collaboratives orientées vers la télé-médecine. Ainsi nous montrons les contraintes plus fortes liées à ce domaine.

1.1/ LE TRAVAIL COLLABORATIF ET LES COLLECTICIELS

Dans cette section nous caractérisons la collaboration et le travail collaboratif, puis nous dégagons un ensemble de concepts génériques de ce domaine.

Les environnements collaboratifs sont basés sur les échanges cohérents des données [LeM04]. Pour bien situer le domaine de la collaboration (Figure 1.1 ①), nous allons tout d'abord définir un domaine plus vaste qui est celui du CSCW (*Computer Supported Cooperative Work* Figure 1.1 ②).

Le terme anglo-saxon CSCW représente un domaine de recherche pluridisciplinaire qui existait depuis les années 90 mais qui a vraiment émergé au début des années 2000. Il s'intéresse au travail en groupe, aux outils et aux supports permettant d'assister ce travail

1. Covalia interactive, <http://www.covalia.com>

de groupe [Jac06]. Il ne s'agit pas seulement de prendre en compte les interactions qu'un individu peut avoir avec un ordinateur mais également les échanges entre les personnes (ou les agents) via un système informatique.

Le CSCW est un domaine pluridisciplinaire dans lequel interviennent, tout d'abord, des recherches provenant des sciences humaines :

- Psychologie :
Cette discipline étudie les répercussions liées à l'utilisation d'un logiciel collaboratif au sein d'un groupe.
- Linguistique :
Ce domaine étudie la sémantique des informations échangées. Par exemple pour des solveurs distribués, la sémantique des données échangées est importante, alors que dans les éditeurs collaboratifs les messages permettent simplement les échanges d'objets qui sont des éléments du contexte partagé.
- Sociologie :
L'étude des comportements sociaux dans le cadre du travail doit être prise en compte dans la réalisation des collecticiels et plus particulièrement les lois qui régissent les comportements de groupe. C'est ainsi que certaines applications collaboratives utilisent la notion de hiérarchie pour gérer la sécurité de l'information. Par exemple une application d'enseignement à distance permettra de séparer les contextes enseignants et élèves.
- Ethnologie :
L'interaction au sein d'un groupe varie suivant les cultures d'où la nécessité d'adapter le collecticiel aux besoins des différentes cultures.

Ensuite, du point de vue de l'informatique le CSCW s'inspire des recherches de domaines variés dont les principaux sont les suivants :

- Le domaine des réseaux et télécommunications :
Les systèmes collaboratifs dépendent souvent directement des capacités des réseaux sous-jacents. Ainsi, le CSCW évolue de pair avec les progrès en réseaux et télécommunications.
- Le domaine des systèmes distribués :
Un logiciel collaboratif est avant tout un système distribué. Il en utilise les différents mécanismes, comme par exemple la mémoire partagée, l'échange de messages ou encore la répartition de charge...
- Nous pouvons citer également les domaines de l'intelligence artificielle et celui des solveurs distribués, mais qui ne sont pas dans nos domaines de recherche...

Les environnements de travail collaboratif (*ETC : appelés aussi collecticiels*) sont des environnements qui permettent de favoriser le travail d'un groupe de personnes [Ben02].

Quatre concepts fondamentaux de ces environnements sont décrits dans la littérature :

- L'individu : qui prend part aux actions de l'environnement de travail, il appartient à un groupe de participants,
- Le groupe : est un ensemble d'individus travaillant dans un même but,
- Le rôle : est défini pour chaque participant au sein du groupe afin qu'il sache quelles actions il doit effectuer selon ses possibilités, ses capacités et ses devoirs



FIGURE 1.2 – Trèfle fonctionnel enrichi

vis-à-vis des autres participants et des données partagées au sein de l'environnement,

- La vue : est propre à chaque participant et adaptée en fonction de ses besoins et des autorisations qu'il détient sur les entités manipulées collectivement,

En s'appuyant sur ces notions fondamentales, les environnements collaboratifs proposent d'enrichir les fonctionnalités habituelles des logiciels par des fonctions qui permettent à un collectif de participants de communiquer pour effectuer des actions de manière coordonnée.

1.1.1/ MODÈLE FONCTIONNEL DES ENVIRONNEMENTS COLLABORATIFS

Afin de pouvoir comparer les différents collecticiels, Ellis définit dans [Ell94] un modèle conceptuel formé de trois dimensions fonctionnelles : la production, la coordination et la communication.

- **L'espace de production** qui caractérise l'activité de l'espace de travail : comme par exemple un tableau blanc pour manipuler des images médicales.
- **L'espace de coordination** qui caractérise les outils, permettant de gérer les interactions des acteurs entre eux et avec le système : par exemple mise en place de protocoles de cohérence ou des protocoles sociaux d'interactions . . .
- **L'espace de communication** qui caractérise tous les échanges, dits informels, entre les acteurs par les canaux audio et vidéo.

Assez rapidement ce modèle a été enrichi dans [Sal95] tout d'abord avec une première évolution vers le modèle des trois C : la Co-production, la Coordination et la Communication. Puis B. David ajoute dans [Dav01] un quatrième espace, celui de la Conversation.

Enfin dans [Elm07] (Figure 1.2) un cinquième espace, celui de la régulation, est orthogonal aux autres. Il caractérise la capacité du système à être régulé ou à s'auto-réguler (ou s'auto-adapter) : comme par exemple la modification des caractéristiques d'une vidéo pour s'adapter à la bande passante disponible, ou la régulation des autorisations d'accès ...

1.1.2/ LES MODES DE FONCTIONNEMENT EN COLLABORATION

En travail collaboratif, il existe plusieurs modes de fonctionnement correspondant à des niveaux d'interaction différents. Les trois principaux modes rencontrés sont :

- **La travail asynchrone** : on parle de coopération et non plus de collaboration,
- **Le travail en session** : dans le monde des systèmes distribués, on parlera de « barrières de synchronisation »,
- **La notion de réunion** : à un instant précis, sur rendez-vous, les acteurs collaborent,
- **La collaboration en temps réel** : c'est la réelle collaboration qui offre un partage *complet* de l'environnement. Les participants interagissent entre eux, et sur l'environnement en temps réel. On retrouve ici les environnements de partage de dessin, d'image, et la télé-conférence.

Le travail collaboratif s'est développé dans différents domaines d'application comme l'apprentissage en ligne ou *e-learning* [Wat10], la maintenance à distance [Ste11], la conception collaborative [Duq09] ou encore la télémédecine [Fui08, Maz14] ...

Dans le cadre de la mobilité des utilisateurs qui devient une nécessité du domaine collaboratif, l'informatique dite ubiquitaire (ou pervasive) a vu le jour [Lou09, Mat03, Dib13] : elle a pour but de rendre accessible une multitude de services depuis n'importe où tout en masquant l'ordinateur. Ainsi depuis les années 2010, l'apparition du *cloud computing* a mis en avant des modes de collaboration particuliers comme le mode SaaS (*Software as a Service*).

1.2/ LE CLOUD COMPUTING

Le cloud computing (l'informatique dans le nuage, Figure 1.3) désigne un ensemble de processus qui consiste à utiliser la puissance de calcul et/ou de stockage de serveurs informatiques distants à travers un réseau, généralement internet.

Il existe une définition proposée par le NIST² (National Institute of Standards and Technology) : « *Le cloud computing est un modèle de technologie qui permet l'accès ubiquitaire, adapté et à la demande sur des ressources configurables partagées* ». En complément de cette définition, le RAD Lab³ de l'Université de Berkeley, explique que le cloud com-

2. NIST, <http://www.nist.gov>

3. RAD Lab, <https://radlab.cs.berkeley.edu>



FIGURE 1.3 – Cloud Computing

puting fait également référence aux applications fournies comme un service sur internet. Les services ont été longtemps connus comme : L'laaS (Infrastructure as a Service), Le PaaS (Platform as a service) et le SaaS (Software as a Service), tandis que le hardware et software dans les centres de données est ce que l'on appelle Cloud. Les trois services d'infrastructure, de plateforme et de logiciel correspondent à des modèles techniques différents. Chacun de ces modèles joue un rôle spécifique [Arm09] :

1. L'laaS correspond à la partie infrastructure du Cloud. Il permet aux entreprises d'externaliser et de faire évoluer leur infrastructure matérielle (serveurs, réseau et stockage) à la demande et à distance. Dans l'laaS, seule l'infrastructure matérielle (hardware) est dématérialisée. C'est par exemple le stockage de données pour des raisons de sauvegardes réalisées par les entreprises qui souhaitent ne pas avoir uniquement des sauvegardes en interne.
2. Le PaaS est un modèle qui s'appuie sur l'laaS. Il permet d'externaliser l'infrastructure matérielle, mais également des applications middleware : bases de données, couches d'intégration de données et environnements de développement des applications.
3. Le SaaS est la couche finale du Cloud, la plus aboutie et la plus simple à appréhender pour l'utilisateur sur web. Le SaaS est un modèle de distribution de logiciels où support logique dans lequel les données manipulées sont hébergées sur des serveurs dans une entreprise de technologie de l'information et de la communication (TIC), qui est accessible, par Internet, via un navigateur Web d'un client. Le fournisseur s'occupe des services de maintenance, de l'exploitation quotidienne et de la maintenance du logiciel utilisé par le client. Régulièrement le logiciel peut être consulté sur n'importe quel ordinateur, présent dans l'entreprise ou non. Il s'ensuit que l'information, le traitement, les entrées et les résultats du logiciel, sont hébergés dans l'entreprise des TIC.

Les principales applications actuelles de ce modèle sont :

- le gestionnaire de relation client (CRM),
- la vidéoconférence,
- la gestion des ressources humaines,
- les communications unifiées,
- et enfin la messagerie et les logiciels collaboratifs.

Grâce à l'utilisation désormais courante du Cloud Computing et aux dernières évolutions technologiques, la télémédecine bénéficie actuellement d'une forte croissance. Les avantages du Cloud dans le domaine de la télémédecine ne sont plus à démontrer : le premier répondant parfaitement aux impératifs de fonctionnement du second [Sul14].

1.3/ TÉLÉ-APPLICATIONS POUR LA MÉDECINE

1.3.1/ UNE CROISSANCE EXPONENTIELLE DEPUIS 2010

Le domaine de la télémédecine émerge fortement (notamment en France depuis la loi HPST : Hôpital, Patient, Santé et Territoires [HPST12] et le décret d'application de novembre 2010). Concernant la France, ce domaine est défini, par la loi, en cinq domaines précis : la téléexpertise, la téléassistance, la téléconsultation, la télésurveillance et la réponse médicale.

D'autre part, le CNOM⁴ définit la télémédecine dans ces termes : *c'est une des formes de coopération dans l'exercice médical, mettant en rapport à distance, grâce aux technologies de l'information et de la communication, un patient (et/ou les données médicales nécessaires) et un ou plusieurs médecins et professionnels de santé, à des fins médicales de diagnostic, de décision, de prise en charge et de traitement dans le respect des règles de la déontologie médicale*. Selon la CNOM, les enjeux prioritaires de la télémédecine sont l'accès équitable aux soins, la maîtrise des dépenses de santé et la qualité des soins.

La télémédecine est utilisée dans le cadre de la surveillance du patient ou pour l'éducation [Wat10] et a également été introduite pour les soins d'urgence. La télémédecine peut être définie par l'utilisation des nouvelles technologies de télécommunication pour fournir des informations et des services médicaux aux patients et aux médecins grâce à la communication visuelle et auditive. Ces nouvelles technologies sont utilisées pour aider les praticiens se trouvant sur des sites distants à effectuer des procédures de diagnostic et de consultation, tels que des examens cliniques grâce aux transferts d'images médicales [Fui08]. Ces images sont souvent volumineuses et peuvent mettre du temps pour être acheminées selon les réseaux utilisés.

Des plateformes collaboratives dans le domaine de la télémédecine ont été mises en place dans les hôpitaux depuis les années 2010. Par exemple, à Besançon, la société Covalia interactive, issue d'un transfert de technologie de notre équipe de recherche, propose d'offrir aux médecins une plateforme coopérative de télédiagnostic appelée CovotemTM (Figure 1.4). Cette plateforme est désormais utilisée entre autres dans les hôpitaux de Basse et haute normandie, les hôpitaux de Paris, les hôpitaux de Martinique

4. Conseil National de l'Ordre des Médecins

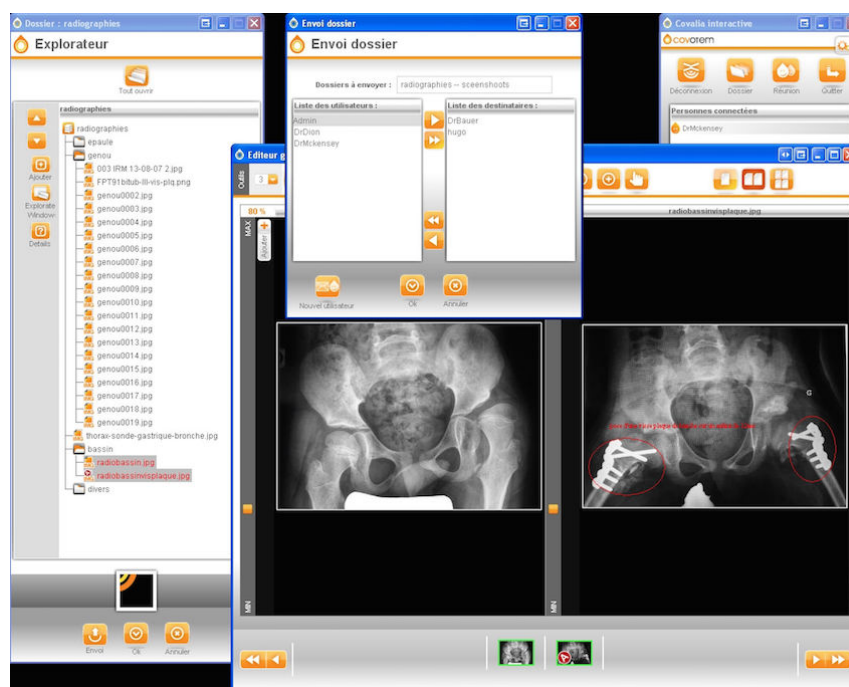


FIGURE 1.4 – Covotem : une plateforme de télédiagnostic temps réel sécurisée

et Guyane . . . Dans les nouveaux outils et plateforme collaboratifs qui sont développés, il existe plusieurs applications pour différentes facettes de la télémédecine :

- La télésurveillance :

La télésurveillance permet de fournir des outils pour surveiller les malades à domicile en utilisant Internet et des réseaux ADSL, GPRS, edge, 3G, . . .

Le projet UR-SAFE [Rum05] propose des solutions de télésurveillance des patients. L'idée est de donner au patient un terminal mobile qui peut récupérer et transmettre des données comme la pression artérielle ou le rythme cardiaque. On peut également citer dans le même domaine le projet EPI-MEDICS [Fay10].

- La téléchirurgie :

En Septembre 2001, le premier cas réel de téléchirurgie a été réalisé par le Professeur Jacques Marescaux [Mar02] spécialiste en chirurgie digestive. Il a opéré depuis New York une patiente qui était à l'hôpital de Strasbourg. Une autre forme de téléchirurgie est la téléassistance permettant d'aider un praticien à effectuer les gestes utiles au diagnostic à distance. Dans ce domaine, un autre projet permet de partager les dossiers médicaux d'un patient en temps réel pour une analyse en profondeur et une préparation optimale des stratégies chirurgicales dans le cadre des tumeurs du foie : c'est le projet Argonaute3D [LeM04] en relation avec France télécom. On peut également citer le robot DaVinci (Figure 1.5) qui est un robot dont la première vocation est ergonomique (position optimal du praticien pour effectuer son geste chirurgical) mais qui propose une distanciation entre chirurgien et champs opératoire qui permet d'envisager la téléchirurgie.



FIGURE 1.5 – Da Vinci : Robot chirurgical avec distanciation

- La téléconsultation :

La téléconsultation est utilisée dans différents domaines de la médecine tels que la neurologie [Gar05] , la psychiatrie, ou en dermatologie pour le dépistage collaboratif des mélanomes [Pie09]. Nous avons cité ici des applications issues de notre équipe de recherche bisontine mais qui, sans être exhaustives, représentent bien les nouveaux usages qui sont désormais courants dans l'utilisation de plateformes collaboratives d'aide au diagnostic médical.

1.3.2/ ORGANISATION COLLABORATIVE DES APPLICATIONS DE SANTÉ

Traditionnellement, les cas cliniques complexes sont traités entre médecins présents physiquement dans un même lieu, principalement pour assurer la confidentialité du patient et une rapide prise de décision. Cependant il est parfois nécessaire qu'une collaboration médicale entre différents professionnels de la santé soit mise en place. Cette collaboration peut se faire au sein de la même institution ou au sein de différentes institutions. Pour cette raison, les services médicaux (les hôpitaux et les centres de santé) ont besoin d'échanger des informations sur les patients, telles que les examens cliniques, les diagnostics et les images exploratoires pour prendre des décisions cliniques et réaliser des diagnostics entre différents professionnels qui ne sont pas nécessairement présents dans le même pays (problème de localisation et d'horaire de travail).

Il peut donc être important de créer une discussion sur un cas clinique avec une interaction ouverte et asynchrone, comme par exemple au travers d'un e-mail, d'un forum ou d'une liste de discussion. Ainsi, l'objectif de mettre à disposition sur une plateforme la description du cas clinique (le dossier du patient et les images exploratoires du patient qui contiennent les diagnostics préliminaires) est de permettre aux autres médecins un accès à ces informations. Ils seront alors en mesure de commenter et de partager leurs

expériences et expertises.

La prise de décision médicale couvre des tâches importantes telles que le diagnostic, la planification du traitement, l'interaction avec les patients, l'identification des erreurs médicales, ... Le diagnostic médical est un processus qui vise à identifier les maladies en se basant sur les symptômes et les rapports de laboratoire. Dans le processus de diagnostic, un schéma de représentation approprié est nécessaire à la fois pour l'interprétation du problème et la récupération de la connaissance. Un système de diagnostic pour les applications de santé a été proposé par Ulieru [Uli04]. Cette application est composée d'une *holarchie* médicale (Figure 1.6), qui est une communauté de personnes et/ou d'entités (hôpitaux, cliniques, bases de données et dispositifs médicaux) ayant pour objectif commun le partage et le contrôle de l'information (par exemple transférer l'information du patient Jean, qui présente une maladie cardiaque). Une plateforme de diagnostic doit être capable de regrouper toutes les ressources implicites dans les diagnostics du patient, tout en ayant pour objectif la surveillance par les médecins de la progression de la maladie grâce à un flux d'informations et à une interaction collaborative.



FIGURE 1.6 – Holarchie médicale

1.3.3/ ASPECT MÉDICO-LÉGAL

Sécurisation des données

Nous avons défini, dans la section précédente, la notion de Cloud Computing qui met à disposition des services informatiques de manière transparente, ainsi ce modèle informatique est parfaitement adapté à la télémédecine. Mais il est nécessaire de vérifier que les technologies sous-jacentes, et la mise en œuvre s'inscrivent dans cadre légal et réglementaire de l'application de la médecine [Pou12]. Par exemple, les données de santé ne peuvent être hébergées que sur un site agréé (*hébergement des données de*

santé à caractère personnel sur support informatique, Article R1111-9 modifié par décret n. 2011-246 du 4 mars 2011 - art. 1.).

L'anonymat et le stockage des données médicales est un aspect important, et des entreprises en France sont agréées (actuellement seulement une petite dizaine d'agréments ont été délivrés), et le simple fait de faire héberger ses données médicales sur leurs sites permet de rendre conforme une application qui utilise des données médicales.

Il faut également ajouter des identifications par *Carte Professionnelle de Santé*, dont la dernière génération est le CPS3 : cette dernière est une carte unique d'identification, d'authentification et de signature électronique, clé d'accès aux différents systèmes d'information. Elle permet de sécuriser l'accès aux données mais également d'identifier quel professionnel de santé est intervenu dans l'acte médical, et quelles responsabilités sont engagées.

Responsabilité de l'acte

Dans ces actes de télémédecine, un autre point très important est celui de la responsabilité de l'acte. En effet, en cas d'erreur médicale il est nécessaire de pouvoir déterminer les responsabilités (aspect juridique, assurance...).

Ainsi pour cet aspect médico-légal, le Conseil d'État (*Rapport public du Conseil d'État : « Réflexions sur le droit de la Santé », 1998*) reconnaît que la recherche de responsabilité en cas de décision diagnostique partagée par plusieurs médecins complique la tâche du juge. Sous réserve de circonstances exceptionnelles, le Conseil d'État estime que le médecin requérant sera seul responsable de la décision diagnostique vis-à-vis du patient, avec la possibilité d'engager une action contre le médecin requis dans le cadre de la responsabilité contractuelle partagée. L'éventuelle faute du médecin requis « *pourra être de nature à dégager totalement ou partiellement le médecin interrogateur (requérant)* ».

Le Prof. Lucas de l'ordre des médecins [Luc12] a ainsi identifié la traçabilité des actes comme indispensable : *le Conseil National de l'Ordre des Médecins affirme que la sécurité physique et informatique des systèmes d'information est une exigence déontologique : cela repose sur l'identification et la traçabilité des accès aux bases de données*.

Ainsi, il en découle que les entreprises qui commercialisent ces produits s'intéressent à un mécanisme qui faciliterait cette recherche de responsabilité : il s'agit de la traçabilité. Actuellement, la seule trace qui reste après un acte de télémédecine, c'est la signature CPS3 de chaque intervenant, mais sans précision sur la responsabilité individuelle, et les actions individuelles des médecins.

Nos travaux s'inscrivent dans cette démarche qui consiste à identifier les actes individuels qui vont concourir à l'élaboration d'un télédiagnostic : tracer quel médecin a effectué quelle action au cours du télédiagnostic.

Dans le chapitre 5 de cette Thèse, nous présentons l'implémentation en *mode SaaS* de notre plateforme : dans cette phase de mise au point qui débouchera sur une phase de tests et validations par les médecins de l'hôpital Civil de Morelia au Mexique, nous n'avons pas encore d'hébergement agréé, ni d'accès par CPS3.

- En phase de recherche, il était difficile de suivre les directives : dossiers volumineux et délais très importants pour obtenir les agréments pour utiliser la CPS3, et coût trop important pour utiliser un serveur agréé,

- Les médecins de l'hôpital civil de Morelia travaillent sur des patients tests qui ne sont pas des patients français et la réglementation mexicaine, pour le moment, est moins contraignante,
- Nous avons sécurisé la connexion à notre site par login/password du médecin,
- Il s'agit d'une utilisation dans le cadre de recherche franco-méxicaine.

CONCLUSION



Depuis les années 2000, un nouveau domaine de recherche a émergé : le domaine du *CSCW* (Computer Supported Collaborative Work). Ce domaine est pluridisciplinaire et il comporte dans son volet informatique, en particulier, l'aspect algorithmique distribuée.

Comme le montre l'étude de la littérature, ces applications distribuées de nouvelle génération nécessitent de respecter certaines contraintes en terme de cohérence des données, de communication, et d'interface distribuée Un nouveau mode de fonctionnement émerge depuis les années 2010 : le cloud computing, qui consiste à utiliser la puissance de calcul et/ou de stockage de serveurs informatiques distants à travers un réseau (le mode *SaaS* est plus particulièrement utilisé).

Enfin, dans le domaine du télédiagnostic médical qui émerge fortement (notamment en France depuis la loi HPST : Hôpital, Patient, Santé et Territoires [HPST12]), les paramètres de sûreté et de sécurité sont encore plus contraignants. Et en particulier, l'aspect médico-légal est primordial. Et il apparaît désormais indispensable de pouvoir tracer les actions effectuées par les différents acteurs d'un acte de télémédecine.

ÉTAT DE L'ART DES ONTOLOGIES MÉDICALES

INTRODUCTION

Les ontologies de domaine sont étudiées et utilisées dans plusieurs domaines de recherche (par exemple la médecine, la biologie, les applications militaires, la philosophie, la linguistique. . .). Les ontologies permettent de stocker des connaissances du domaine dans un formalisme partagé. A l'inverse des bases de données, elles offrent la possibilité de stocker la sémantique d'un domaine et de fournir des mécanismes d'inférence.

Au cours des dernières décennies, la recherche dans le domaine de l'informatique biomédicale s'est développée. Dans le même temps, divers outils biomédicaux sont conçus pour exécuter la collecte et la gestion des données. Aujourd'hui, les systèmes de base de données sont confrontés à la gestion d'une grande quantité d'information (par exemple dossiers cliniques, diagnostics des patients, examens de laboratoire, images médicales. . .). Il est donc très important :

- (1) de collecter et stocker correctement les informations médicales produites au cours des consultations, traitements et études médicales ;
- et (2) de collecter des connaissances utiles à partir de ces données, pour fournir le meilleur traitement possible aux patients.

Ces efforts visent à créer des modèles qui permettent de formaliser d'une manière efficace toutes les connaissances issues des activités médicales de sorte qu'elles puissent être utilisées par la suite, et de rendre plus efficace la détection et le traitement des pathologies connues. Parmi les mécanismes de formalisation, largement utilisés ces dernières années, on trouve les *ontologies*. Elles s'intéressent à la création d'entités : les propriétés et les relations de termes concrets ou abstraits. L'ontologie s'appuie sur la mise en place d'un vocabulaire commun et utilise des représentations et des concepts définis par des ressources terminologiques (par exemple UMLS¹, MeSH², SNOMED³. . .) et des ressources ontologiques construites pour répondre à des besoins précis mais divers (par exemple OBO⁴ qui est une bibliothèque d'ontologies formelles), elles sont expliquées dans la section 2.2.1. Ces ressources permettent l'interopérabilité des données et facilitent l'ajout de connaissances.

1. UMLS <http://www.nlm.nih.gov/research/umls/>

2. MeSH <http://www.ncbi.nlm.nih.gov/mesh>

3. SNOMED <http://www.ihtsdo.org/snomed-ct/>

4. OBO <http://www.obofoundry.org>

Dans chaque pays ou continent apparaissent des groupes multidisciplinaires qui cherchent à standardiser les connaissances et le langage médical afin de mieux utiliser ces ressources.

Dans ce chapitre, nous présentons tout d'abord le domaine général de recherche sur les ontologies, puis les spécificités des ontologies développées et utilisées dans le cadre médical. Enfin les dernières sections de ce chapitre présentent deux points essentiels qui sont le cycle de vie d'une ontologie et son évaluation.

2.1/ LE DOMAINE GÉNÉRAL DES ONTOLOGIES

La notion d'ontologie s'intéresse à la fois à l'ingénierie des connaissances, à la linguistique et à la philosophie. Initialement, l'ontologie était un domaine de la philosophie concernant « *l'étude de l'être en tant qu'être, c'est-à-dire l'étude des propriétés générales de ce qui existe* ». En philosophie, on peut parler d'une ontologie comme d'une théorie de la nature de l'existence. En informatique et sciences de l'information, l'ontologie est un terme technique désignant un artefact qui permet la modélisation de connaissances sur un domaine quelconque qu'il soit réel ou imaginaire.

Le terme a été ensuite adopté en Intelligence Artificielle (IA). Les chercheurs en IA ont reconnu l'applicabilité du travail de la logique mathématique et ont fait valoir qu'il est possible de créer de nouvelles ontologies comme des modèles informatiques qui permettent un raisonnement automatisé [Mac89]. Dans les années 1980, la communauté IA a utilisé le terme ontologie pour désigner à la fois une théorie d'un monde modélisé et une composante des systèmes de connaissances [Gru93a]. Nous pouvons décrire l'ontologie d'un programme comme un ensemble de termes. Dans une telle ontologie, les définitions de termes associent les noms des entités de l'univers du discours avec des textes compréhensibles par les humains qui décrivent la signification des concepts et des axiomes formels qui limitent l'interprétation et la bonne utilisation de ces termes.

Il existe de nombreuses définitions de l'ontologie. Malgré leurs similarités, elles diffèrent sur certains points. La définition la plus largement acceptée est celle donnée par Gruber [Gru93a] : « *une ontologie est une spécification explicite d'une conceptualisation* ». Une conceptualisation est un résumé, une vue simplifiée du monde, que nous souhaitons représenter pour une certaine intention. Cette vue est composée d'objets, de concepts, et d'autres entités qui sont censés exister dans un domaine d'intérêt ainsi que des relations qui existent entre eux [Gru93b]. Les définitions d'ontologie ont provoqué beaucoup de débats, les points les plus communément retenus sont :

- L'ontologie définit des concepts, des relations et d'autres distinctions qui sont pertinents pour la modélisation d'un domaine [Gru95].
- La spécification prend la forme de définitions du vocabulaire de représentation (classes, relations...), qui fournissent des significations pour le vocabulaire et les contraintes formelles sur son utilisation cohérente [Gua98].

Les ontologies partagent des propriétés avec les taxonomies [Hak04] (liste de termes contrôlés organisés de façon hiérarchique), les classifications et également les thésaurus (réseaux de termes contrôlés, enrichis par des relations associatives) [Vii04]. Pour spécifier une conceptualisation, il est nécessaire de définir les axiomes qui peuvent contraindre l'interprétation des termes définis. Charlet [Char02] propose quant à lui cette

définition : « *une ontologie implique ou comprend une certaine vue du monde par rapport à un domaine donné* ». Cette vue est souvent conçue comme un ensemble de concepts (entités, attributs, processus), de définitions et d'interrelations. Gandon [Gan06] enrichit cette définition en indiquant qu'une ontologie informatique est une représentation de propriétés générales de ce qui existe, que l'on peut formaliser et qui peut supporter un traitement rationnel.

Le développement d'ontologies s'est effectué dans des domaines aussi variés que la médecine, le droit, la biochimie, l'indexation de séquences audiovisuelles, l'électronique... Les ontologies apparaissent aujourd'hui comme des composants logiciels avancés qui s'insèrent au centre des systèmes informatiques pour leur apporter une dimension sémantique [Bod06a].

Après plusieurs critiques et suggestions, en 2009, Gruber propose une nouvelle définition de l'ontologie [Gru09] : « *Dans le contexte de l'informatique et des sciences de l'information, une ontologie définit un ensemble de primitives de représentation qui permettent la modélisation d'un domaine de connaissance* ». Les primitives de représentation sont généralement des classes (ou ensemble), des attributs (ou propriétés), et des relations (ou des relations entre les membres de la classe). La définition de primitive de représentation comprend des informations sur leur signification et les contraintes sur leur application logique et cohérente. Dans le contexte des systèmes de base de données, l'ontologie peut être considérée comme un niveau d'abstraction des modèles de données, analogues à des modèles hiérarchiques et relationnels, mais conçue pour la modélisation des connaissances sur les individus, leurs attributs et leurs relations avec d'autres individus. Les ontologies sont généralement spécifiées dans des langages qui permettent l'abstraction à partir des structures de données et du contexte d'application.

En pratique, les langages des ontologies sont plus proches de l'expressivité de la logique du premier ordre que les langages utilisés pour modéliser des bases de données. Pour cette raison, les ontologies sont censées être au niveau *sémantique*, alors que le schéma de base de données est un modèle de données au niveau *logique* ou au niveau *physique*. Dû à leur indépendance, les ontologies sont utilisées pour intégrer des bases de données hétérogènes, permettant l'interopérabilité entre des systèmes disparates, et en spécifiant des interfaces de services indépendants, basées sur la connaissance. Dans la technologie des normes du Web sémantique [Ber01], les ontologies sont composées de plusieurs couches (*cf.* section 2.1.3). Il y a maintenant des langages standards et une variété d'outils commerciaux et *open source* pour créer et travailler sur des ontologies.

Dans une perspective plus pragmatique, on peut dire que l'ontologie est un outil qui produit de l'ingénierie des connaissances dépendante de son domaine d'application (*cf.* section 2.1.1). Ainsi, les ontologies fournissent des mécanismes pour instancier des modèles d'un domaine dans des bases de connaissances, faire des requêtes basées sur ces connaissances et représenter les résultats obtenus. Malgré tout, la création d'une ontologie constitue une tâche complexe pour le développeur. C'est pourquoi, un ensemble de méthodologies ont été mises en œuvre pour faciliter cette tâche (*cf.* section 2.1.2). La représentation de l'ontologie s'appuie quant à elle sur un formalisme spécifique qui a été mis en avant par le *W3C-Semantic Web*. Ce formalisme permet de coder des ontologies et possède plusieurs variantes avec des expressivités différentes (*cf.* section 2.1.3). Cela reflète l'intention qu'une ontologie est une spécification d'un modèle abstrait de données (la conceptualisation du domaine) indépendant d'une structure particulière. Une ontologie définit un vocabulaire qui permet la formulation d'assertions. De même, si

une ontologie doit être formulée dans un langage de représentation, il est destiné à être une spécification de niveau sémantique, c'est-à-dire, qu'il est indépendant de la stratégie de modélisation de données ou de la mise en application.

2.1.1/ TYPES D'ONTOLOGIES

Comme dans le cas de la définition d'une ontologie, il n'existe pas non plus de consensus à propos de la classification des différents types d'ontologie. Valencia [Val05] montre une synthèse des diverses classifications existantes. La Figure 2.1 illustre cette classification.

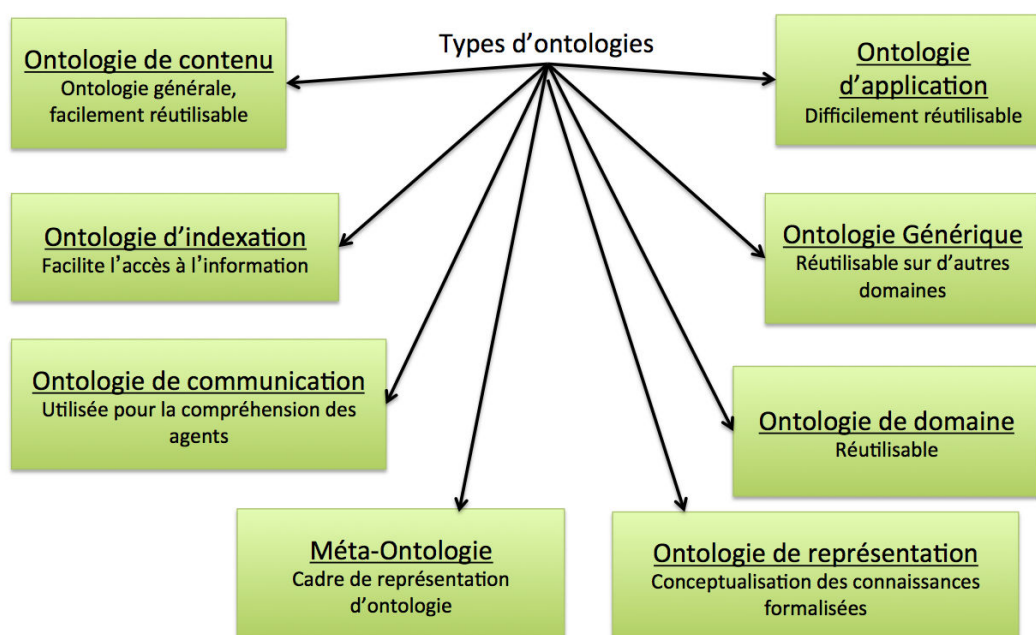


FIGURE 2.1 – Classification des ontologies pour le type de problème à résoudre

Kerpler [Ker06] détaille une autre classification conçue pour les ontologies médicales. Mais cette classification est également applicable à d'autres types d'ontologie tels que :

- les ontologies linguistiques (construites par des vocabulaires contrôlés, des taxonomies, des thésaurus et des glossaires),
- les ontologies destinées pour une application (constituées de schémas conceptuels et de bases de connaissances)
- et les ontologies formelles (construites par des ontologies de domaine, des ontologies de bases ou de références et des ontologies fonctionnelles).

Ce même auteur souligne que les catalogues et les schémas conceptuels ne doivent pas être considérés comme des ontologies.

Tout objet contenant des connaissances peut être considéré comme une ontologie [Gru09]. Malgré tout, certains éléments sont nécessaires afin de constituer/construire une ontologie. Ces éléments sont les suivants :

- les **Classes** : elles sont également appelées concepts ou termes. Elles correspondent aux abstractions pertinentes d'un segment de la réalité (le domaine du problème) retenues en fonction des objectifs qu'on se donne et de l'application envisagée pour l'ontologie. Habituellement, elles sont organisées en taxonomies. Elles ne doivent pas

être considérées à elles seules comme une ontologie complète, mais elles peuvent être considérées comme un thésaurus.

- les **Caractéristiques** : elles constituent la structure interne des classes et peuvent être propres ou héritées lorsque la classe agit comme fille d'une autre classe, en raison de la taxonomie.
- les **Relations** : elles traduisent les associations (pertinentes) existantes entre les concepts présents dans le segment analysé de la réalité. Ces relations incluent les associations : « sous_classe_de », « partie_de », « associe_à », « instance_de », ... Ces relations nous permettent d'apercevoir la structuration et l'interrelation des concepts, les uns par rapport aux autres.
- les **Fonctions** : elles constituent des cas particuliers de relations, dans lesquelles un élément de la relation, (le N^{ème}) est défini en fonction des N-1 éléments précédents.
- les **Axiomes** : ils constituent des assertions acceptées comme vraies, à propos des abstractions du domaine, définies par l'ontologie. Ils sont utilisés pour définir le sens des composants ontologiques, définir des contraintes complexes sur les valeurs des attributs, les arguments des relations et les concepts. Il est nécessaire de vérifier la complétude des informations spécifiées dans l'ontologie ou dans le cas contraire de décider de nouvelles informations à introduire. Il est également possible d'utiliser la logique du premier ordre pour les décrire.
- les **Instances** : elles permettent d'enrichir l'ontologie, ces objets véhiculent les connaissances (statiques, factuelles) à propos du domaine du problème.

2.1.2/ LES MÉTHODOLOGIES DE DÉVELOPPEMENT DES ONTOLOGIES

Le développement d'une ontologie en partant de zéro ou en réutilisant une ontologie existante n'est pas une tâche facile. Pour cette raison, des méthodologies ont été mises en place pour faciliter ce développement. Ces méthodologies décrivent les différentes étapes à suivre pour obtenir de manière rapide (et sûre) un ensemble de connaissances de qualité issues de sources multiples. L'absence de standard et le nombre important de méthodologies ralentissent le développement d'ontologies et rendent plus complexe le travail des développeurs. Ces méthodologies permettent également la réutilisation et l'extension d'ontologies existantes.

On entend par méthodologie : les procédures de travail, les étapes qui décrivent le pourquoi et le comment de la conceptualisation puis l'artefact construit. Ce qui suit présente les méthodologies les plus référencées.

La méthodologie Cyc [Len90] consiste à établir manuellement la connaissance commune qui est implicite dans différentes sources. Une fois que l'on possède suffisamment de connaissances dans notre ontologie, on peut acquérir de nouvelles connaissances en utilisant des outils classiques de traitement du langage naturel.

La méthodologie d'Uschold et King [Ush95] propose trois étapes générales pour le développement des ontologies : (1) identifier le but de l'ontologie ; (2) capturer les concepts et relations entre ces concepts et les termes utilisés ; (3) codifier l'ontologie. L'ontologie doit être documentée et évaluée, il est possible d'utiliser d'autres ontologies pour en créer une nouvelle.

La méthodologie présentée par Grüninger et Fox [Gru95] est constituée de deux étapes : (1) identifier intuitivement les applications possibles pour utiliser une ontologie ; et (2) générer une série de questions en langage naturel (appelée questions de compétence) pour déterminer l'apport de l'ontologie. Ces questions sont utilisées pour extraire les principaux concepts, les propriétés, les relations et les axiomes, qui sont définis formellement en Prolog. En conséquence, il s'agit d'une méthodologie formelle qui s'appuie sur la robustesse de la logique classique et qui peut être utilisée comme guide pour transformer des scénarios informels dans des modèles informatiques. Cette méthodologie a été utilisée pour la construction de l'ontologie TOVE⁵.

La méthodologie « *Ontology development 101* » [Noy01] proposée par l'Université de Stanford possède les principales recommandations suivantes : (1) déterminer le domaine et l'apport de l'ontologie (avec quelques questions de base : quel est le domaine que va couvrir l'ontologie ? Dans quel but utiliserons-nous l'ontologie ? Qui va utiliser et maintenir l'ontologie ? ...) ; (2) déterminer l'intention d'utilisation de l'ontologie ; (3) envisager une éventuelle réutilisation des ontologies existantes ou vocabulaires existants ; (4) énumérer les termes importants du domaine ; (5) définir les classes et la hiérarchie des classes ; (6) définir les propriétés des classes et attributs ; (7) définir les facettes des attributs ; et (8) créer les instances.

Terminae⁶ [Bie99] apporte à la fois une méthodologie comme un outil d'ingénierie des connaissances assisté par ordinateur écrit en Java. L'originalité de cet outil est d'intégrer des outils linguistiques et de l'ingénierie des connaissances (l'utilisation des relations linguistiques entre les notions et leurs synonymes). Il aide à représenter une notion comme un concept, lequel est appelé le concept terminologique. Terminae propose une méthodologie pour construire des concepts terminologiques à partir de l'étude des termes correspondants dans un corpus. Terminae est composé des étapes suivantes : (1) établir la liste des termes qui constituent un corpus de textes pertinents sur le domaine, (2) LEXTER (un extracteur de termes) [Aus00], cet outil propose à l'ingénieur un ensemble de termes candidats qui doivent être choisis avec l'aide d'un expert, (3) conceptualiser chaque terme (l'ingénieur analyse les usages du terme dans le corpus pour définir toutes les notions du terme), (4) définir en langage naturel chaque notion, et (5) traduire la notion dans un formalisme (par exemple, logique descriptive, TBox, ABox).

OntoClean [Gua09] est une méthodologie de validation de l'adéquation ontologique des relations taxonomiques. Elle est basée sur les notions ontologiques générales tirées de la philosophie (l'essence, l'identité et l'unité), et est utilisée pour caractériser des aspects importants de la signification voulue des propriétés, des classes et des relations qui font une ontologie. Ces aspects sont représentés par des méta-propriétés formelles, qui imposent plusieurs restrictions sur la structure taxonomique d'une ontologie. L'analyse de ces restrictions aide à l'évaluation et à la validation des choix effectués.

5. TOVE Ontology, <http://www.eil.utoronto.ca/enterprise-modelling/tove/>

6. Terminae, <http://lipn.univ-paris13.fr/fr/article/1314-terminae/166>

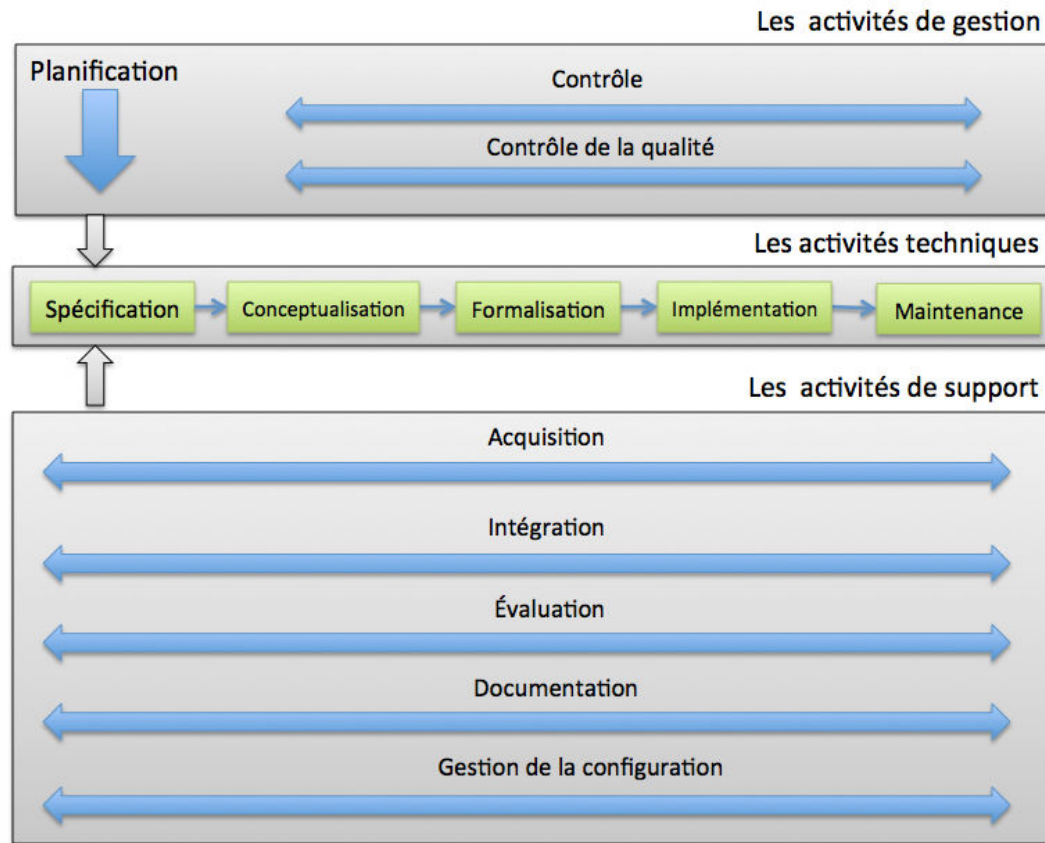


FIGURE 2.2 – Processus de développement d'ontologie de Methontology

Methontology⁷ [Cor03] développée au laboratoire d'intelligence artificielle de l'Université Polytechnique de Madrid. C'est une méthodologie qui s'applique dans le cadre du développement d'ontologies à partir de rien, de la réutilisation d'ontologies ou d'un processus de ré-ingénierie. Cette méthodologie propose trois principes : (1) une identification du processus de développement de l'ontologie au travers des actions suivantes : évaluation, gestion de configuration, conceptualisation, formalisation, intégration, implémentation, et maintenance ; (2) un cycle de vie basé sur des prototypes évolués ; et (3) une méthodologie déjà mentionnée, qui spécifie les étapes à exécuter dans chaque activité, les techniques utilisées, les produits à obtenir et comment ils doivent être évalués [Cor05]. La Fondation pour les agents physiques intelligents (FIPA - Foundation for Intelligent Physical Agents) qui promeut l'interopérabilité entre les applications basées sur des agents [Gom04, Cor05] recommande d'utiliser Methontology, comme méthodologie pour le développement des ontologies. Le cycle de vie de cette méthodologie est décrit dans la Figure 2.2. Les activités de contrôle, l'assurance de la qualité, l'acquisition de connaissances, l'intégration, l'évaluation, la documentation et la gestion de configuration sont effectuées simultanément avec les activités de développement.

La méthodologie NeOn⁸ est un projet européen, qui a débuté en Mars 2006 avec la participation de chercheurs de 14 pays européens, dans le domaine de l'ingénierie on-

7. Methontology, <http://semanticweb.org/wiki/METHONTOLOGY>

8. NeOn, <http://www.neon-project.org>

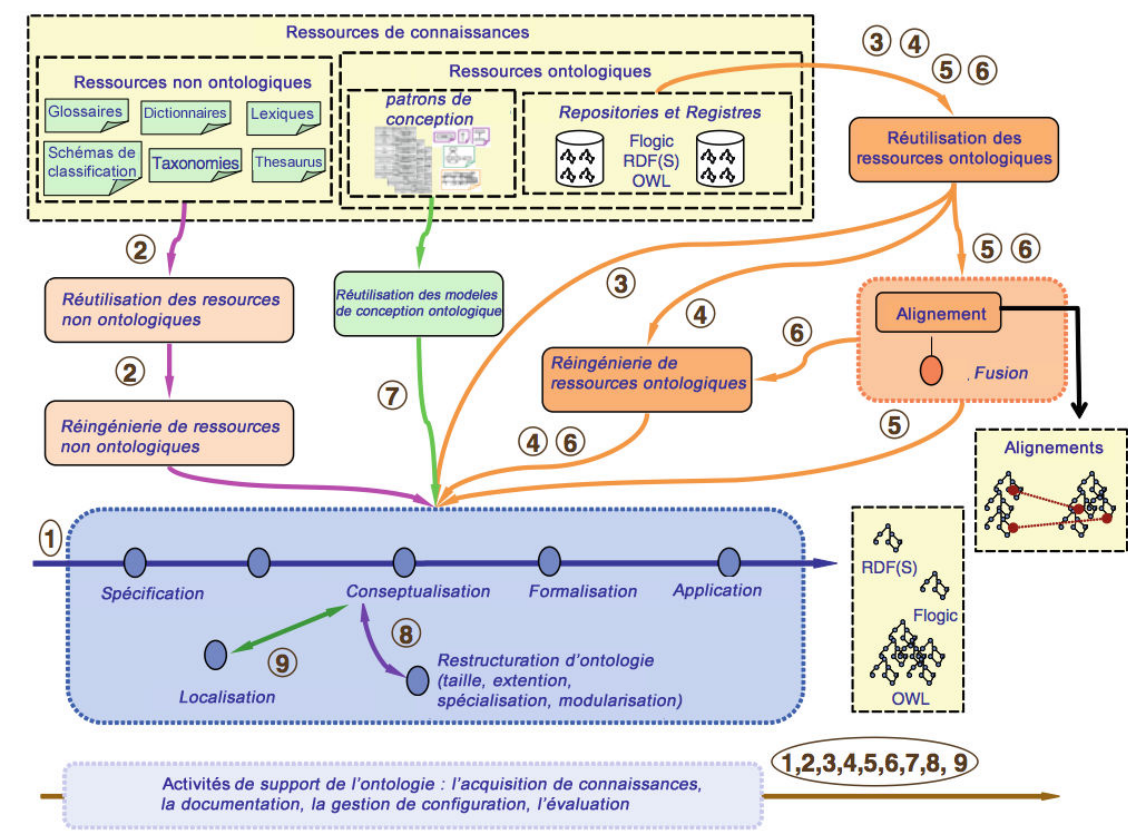


FIGURE 2.3 – Processus de développement d'ontologie de NeOn [Sua12]

tologique, des technologies de collaboration, de l'ingénierie logicielle et de l'interaction homme-machine [Sua12]. Son objectif principal est d'améliorer la capacité à gérer plusieurs ontologies sur le réseau (y compris l'évolution et la maintenance de celles-ci), ainsi qu'à guider le processus de développement des ontologies collaboratives et à réutiliser des ressources ontologiques et non-ontologiques. En ce sens, les auteurs du projet ont souligné que les ontologies sur le réseau ne sont pas des artefacts indépendants. Elles sont sémantiquement reliées entre elles par les relations suivantes (Figure 2.3) :

- les importations et les dépendances. L'une des relations des plus courantes entre deux ontologies est la «dépendance» : une ontologie nécessite la soumission des définitions incluses dans une autre ontologie,
- le contrôle de version. L'activité de suivi des différentes versions d'une ontologie et le processus d'évolution de l'ontologie doivent être surveillés, afin d'assurer que l'information est mise à jour,
- les alignements. Ils consistent en la correspondance (mapping) de ces différents modèles, en précisant quelles entités doivent être considérées comme égales ou quand une est plus générale qu'une autre,
- la modulation. Elle permet l'identification des composants de l'ontologie qui peuvent être considérés séparément et qui participent chacun à un sous-domaine.

La méthodologie propose également un cadre de travail pour le développement de réseaux des ontologies qui comprend neuf scénarios : (1) la spécification d'application, (2) la réutilisation et la ré-ingénierie des ressources non-ontologiques (NOR), (3) la réutilisation des ressources ontologiques, (4) la réutilisation des ressources d'ingénierie

ontologique, (5) la réutilisation des ressources et la fusion ontologique (6) la réutilisation, la fusion et la ré-ingénierie des ressources ontologiques, (7) la réutilisation des modèles de conception pour l'ontologie (PDOs), (8) la restructuration des ressources ontologiques, (9) la localisation des ressources ontologiques.

Nous présentons les méthodologies les plus utilisées. Il en existe un grand nombre non représentées ici, car ce n'est pas le but de cette recherche. Nous détaillons les plus importantes les plus référencées et utilisées pour la construction des ontologies. Pour conclure, on peut dire que le développement des ontologies doit tenir compte de la relation avec l'architecture du système d'information, sans oublier l'importance des formulations théoriques de la connaissance sur un domaine particulier.

2.1.3/ LES LANGAGES DE REPRÉSENTATION

Plusieurs propositions sont apparues sur la manière la plus appropriée de représenter les ontologies dans un langage compréhensible par la machine et par l'homme en même temps. Actuellement les langages RDF et OWL sont les plus utilisés [Tho11]. La Figure 2.4 montre une évolution de ces langages.

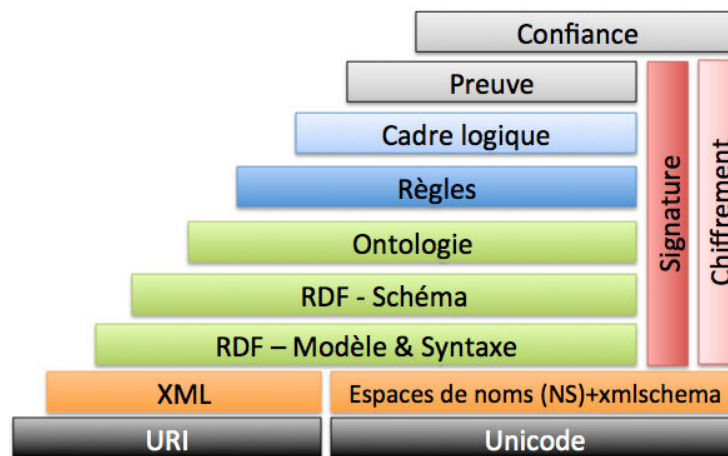


FIGURE 2.4 – L'architecture de niveaux de Berners Lee

Au début des années 90, les langages les plus représentatifs étaient Cycl créé en 1990 par Lenat et Guha [Len90], LOOM créé en 1991 par McGregor [Mac91], Ontolingua créé par Gruber en 1992 [Gru92], F-logic créé par Kifer en 1995 [Kif97], OCML créé par Motta en 1998 [Mot98, Mot00]. La plupart de ces langages suivent une syntaxe basée sur LISP et sont appelés « *langages classiques* ». Plus récemment, l'évolution de ces langages s'est appuyée sur un langage largement utilisé dans le web (XML - *Extensible Markup Language*) pour donner de nouveaux langages pour les ontologies comme OML/CKML [Ken00], RDF⁹ (*Ressource Description Framework*), RDF Schema¹⁰, SHOE¹¹, OIL, DAMIL + OIL [Hor02], ou OWL¹² (*Ontologies Web language*).

Les langages de description et de représentation des connaissances évoluent dans

9. RDF <http://www.w3.org/2001/sw/wiki/RDF>

10. RDF Schema <http://www.w3.org/TR/PR-rdf-schema>

11. SHOE <http://www.cs.umd.edu/projects/plus/SHOE>

12. W3C, OWL <http://www.w3.org/2001/sw/wiki/OWL>

tous les domaines. Ainsi, la connaissance du domaine, en particulier les ontologies, est représentée en utilisant le formalisme des graphes conceptuels ou les logiques de description. Depuis l'avènement du *web sémantique*, la représentation des ontologies est en passe d'être normalisée par un langage : OWL qui intègre nativement des fonctionnalités telles que la création d'axiomes qui concernent la définition complète ou partielle de concepts et de relations. Ce langage permet de manipuler et de spécifier des classes, les propriétés de ces classes et des restrictions sur ces propriétés (OWL-DL et OWL-Full deux variantes de OWL). OWL est basé sur une sur-couche XML et sur les logiques de description provenant des graphes conceptuels [Gan02].

Par rapport à la problématique de la représentation des connaissances issues de textes, XML devient le standard de fait de l'ingénierie documentaire textuel [W3C13]. De plus, il fournit avec une sur-couche comme RDF [Wei09], une façon de décrire des méta-données permettant l'indexation et l'accès à l'information [Cru04].

Les technologies, langages, outils et standards du *Web Sémantique* (XML, RDF, OWL) jouent également aujourd'hui un rôle déterminant dans tout type d'applications de domaine qui ont des besoins réels et concrets d'utilisation du web. Des enjeux importants sont de savoir si les langages standards d'ontologies du web permettront de répondre à ces principaux besoins [Gan02, Gan06].

2.2/ LES ONTOLOGIES DANS LE DOMAINE MÉDICAL

Une ontologie médicale est souvent conçue pour assister l'analyse qualitative des décisions médicales dans un domaine spécialisé. La modélisation s'appuie sur deux sources complémentaires [Sad12] :

- la structure de l'observation médicale informatisée utilisée dans l'unité qui fournit directement des concepts fondamentaux pour la prise en charge d'une maladie ou le suivi hospitalier (diagnostics des patients, dossiers médicaux, médicaments administrés, ...),
- l'analyse terminologique des corpus qui permet d'enrichir et de structurer ce noyau ontologique.

La construction de ces types d'ontologie médicale pour la recherche d'information est problématique au niveau de la terminologie et de la modélisation (à la croisée de plusieurs domaines tels que l'ingénierie des connaissances et la gestion pour l'optimisation des coûts en médecine) pour permettre l'interopérabilité des documents et faciliter l'élaboration de diagnostics plus précis [Sta09].

Le développement d'une ontologie utilise obligatoirement des systèmes terminologiques, bien que les ontologies aient un rôle normatif analogue aux terminologies. Pourtant, le recours à des terminologies ou des corpus de textes, pour la création d'ontologies est parfois inévitable (par exemple si l'ontologie a pour intérêt à être intégrée à un système de traitement automatique de documents), l'ontologie doit assurer la couverture terminologique du domaine. Depuis de nombreuses années, l'accès aux connaissances médicales est un enjeu majeur pour les professionnels de la santé et les chercheurs. Face à la multiplication des sources d'information potentiellement accessibles et à l'augmentation vertigineuse de la production textuelle électronique et d'imagerie l'utilisation des ontologies s'est développée.

Plusieurs communautés d'ingénierie de connaissances travaillent depuis plus d'une dizaine d'années sur le problème de la construction de ressources terminologiques et ontologiques à partir de corpus. Ils ont produit des résultats tant théoriques que méthodologiques et logiciels, qui ont été éprouvés dans plusieurs projets applicatifs. Depuis 2000, ces projets sont reconnus comme étant des avancées sur ces problématiques au niveau international [Aus05]. Le développement de ces ressources terminologiques et ontologiques facilite l'usage des terminologies nationales et internationales disponibles notamment dans le domaine de la médecine (aide au codage des diagnostics, réalisation d'études épidémiologiques et maladies, ...) et pour l'accès aux connaissances médicales (base de données bibliographiques MEDLINE¹³, base de connaissance VIDAL¹⁴ sur les médicaments, documents disponibles sur internet, ...). Les utilisateurs potentiels de ce type de ressources sont nombreux : les étudiants et les médecins, les organismes et institutions publics ou privées, les services dans les hôpitaux, Dans le contexte du *web sémantique*, l'interopérabilité des systèmes d'information de santé passe par la constitution de ressources terminologiques et ontologiques [Char02, Ban05].

La quantité de connaissance disponible dans les domaines de la médecine, de la biologie et de la santé publique s'accroît [Saf00]. Les documents médicaux numérisés constituent des sources multiples et diverses (texte, vidéo, audio, images, ...). Ces numérisations font espérer des bénéfices [Ash04] :

- l'augmentation de la fiabilité des données (capture, enregistrement, transmission),
- la compréhension des mécanismes d'interprétation et de raisonnement médical,
- la sélection des données,
- la rationalisation des choix au niveau individuel ou collectif par l'application de protocoles,
- le partage de l'information,
- et la facilitation de l'accès à la connaissance [Van11].

De la même façon, l'informatique médicale s'attache à développer et à évaluer des méthodes et des systèmes pour l'acquisition, le traitement et l'interprétation des données (étude des patients), avec l'aide des connaissances issues de la recherche scientifique.

La prolifération des recherches scientifiques ne permet pas de réaliser une carte exhaustive des équipes travaillant sur ce domaine médical. Pourtant, les nécessités de l'activité médicale avec des dossiers informatisés, les nécessités d'indexation, les contraintes de codage, ... obligent à réfléchir à des modélisations puis à des applications qui correspondent à des prises en charge de l'information et des connaissances différentes. En particulier, la diversité de l'activité implique de s'intéresser, d'une part à la question des documents et de leurs supports et des modèles d'information non structurés, et d'autre part à des modèles formels des connaissances que sont les ontologies.

Des nombreux travaux de recherche visant à proposer de nouvelles méthodes de constructions d'ontologies, vont dans ce sens et enrichissent la définition de la notion d'ontologie médicale. Il existe différents centres de recherche ontologiques rattachées aux Universités françaises et au niveau international. Pour exemple, NCOR est le centre national de la recherche ontologique (*National Center for Ontological Research, University of Buffalo*), et a été créé en 2005 avec l'objectif de faire progresser la qualité de la recherche et du développement ontologique et l'assurance de qualité. De la même façon,

13. MEDLINE <http://www.medline.com>

14. VIDAL <http://www.vidal.fr>

le centre national d'ontologie biomédicale (*National Center for Biomedical Ontology*) qui est maintenant dans sa septième année, et HeTOP (*The health Terminology/Ontology Portal*) développé par Stefan Darmoni, sont des centres qui visent (1) à créer et maintenir un référentiel d'ontologies et de terminologies biomédicales, (2) à construire des outils et des services web pour permettre l'utilisation des ontologies et terminologies dans la recherche clinique et translationnelle (disponibles principalement en français ou en anglais, mais aussi en allemand, italien, chinois, ...). La pièce maîtresse du centre national est une ressource en ligne continue sous le nom BioPortail, et HeTOP peut être utilisé par les humains et les ordinateurs via des services Web.

Au niveau national, le Groupe d'Informatique Biomédicale (GIB) a travaillé durant 15 ans dans divers domaines de l'informatique biomédicale sur entre autres : la prise de décision, le traitement d'images, les lignes directrices et protocoles cliniques, les ontologies biomédicales, l'intégration de bases de données, le data-mining, ... L'expérience acquise dans ces domaines se reflète dans le grand nombre de publications scientifiques, dans la participation à différents projets de la Commission Européenne et dans les investissements faits par les industriels. Le groupe a participé à de nombreux projets financés par la Commission européenne, comme ACGT (essais cliniques post-genomique axés sur le cancer) ; INFOBIOMED (Réseau d'excellence en informatique biomédicale) ; INFOGENMED (développement d'un laboratoire virtuel pour l'accès et l'intégration des informations génétiques médicales et les applications dans le domaine de la santé) et BIOINFOMED (en mettant l'accent sur les relations et les synergies entre l'informatique médicale et bio-informatique). En 2011, un projet d'action de soutien de type (ACTION-GRID) a été négocié avec la Commission Européenne pour étendre les applications de la grille de la biomédecine pour les pays d'Amérique Latine, les Balkans et l'Afrique du Nord.

Pour le développement d'une ontologie médicale, les ressources terminologiques sont utilisées par les professionnels de la santé dans le monde entier, dans le cadre de la communication des connaissances et des pratiques scientifiques, s'appuyant sur des ressources documentaires (*cf.* section 2.2.1). Ces ressources aident à la construction d'ontologies dans le domaine biomédical. Elles sont basées sur un traitement automatisé de l'information de façon intensive, standardisée et généralisée pour la représentation des connaissances (*cf.* section 2.2.2). La section 2.2.3 introduit les ontologies médicales les plus reconnues et utilisées dans le domaine de la pratique médicale. Finalement, nous considérons certaines des tendances en matière de développement et traitement ontologique (*cf.* section 2.3) ainsi que l'évaluation d'une ressource terminologique et ontologique (*cf.* section 2.4).

2.2.1/ TAXONOMIES MÉDICALES

Il existe dans le domaine médical un grand nombre de ressources terminologiques et ontologiques construites pour répondre à des besoins précis et divers. L'utilisation de systèmes terminologiques pour la création d'ontologies ne va pas sans poser d'importants problèmes. Bien que les ontologies aient un rôle normatif analogue aux terminologies, mettre en place un vocabulaire commun et faire usage de représentations et concepts partagés, permet l'interopérabilité des documents et facilite l'élaboration de connaissances.

Les ressources disponibles dans le domaine biomédical peuvent être classées comme

des lexiques, des bases de données ou des ontologies. Les lexiques sont constitués d'une collection de mots enrichie d'informations pour leur compréhension et leur utilisation [Sch12]. Nous pouvons rencontrer aussi des *terminologies*, qui sont généralement considérées comme des *lexiques spécialisés* [Bod06b]. Les bases de données se fondent sur le traditionnel modèle relationnel qui fournit un schéma strict pour des instances. D'autre part, les concepts sont représentés par des tables et les relations par des clés étrangères entre les tables. Ce formalisme limite les définitions formelles possibles des concepts et n'offre pas beaucoup de capacités de raisonnement, même s'il existe des approches pour stocker des ontologies dans des bases de données [Lep08]. Les ontologies de domaine ont des objectifs plus précis, alors que leurs administrateurs sont des applications informatiques plutôt que des personnes. Ainsi, les ontologies de domaine n'ont pas besoin de s'occuper des variantes et des catégories syntaxiques sur les termes qui sont utilisés parce qu'elles sont généralement modélisées par un langage de représentation. Nous identifions les formalismes simples comme les réseaux sémantiques et les langages de représentation complexes qui permettent d'appliquer l'inférence comme la structure logique ou la logique de description.

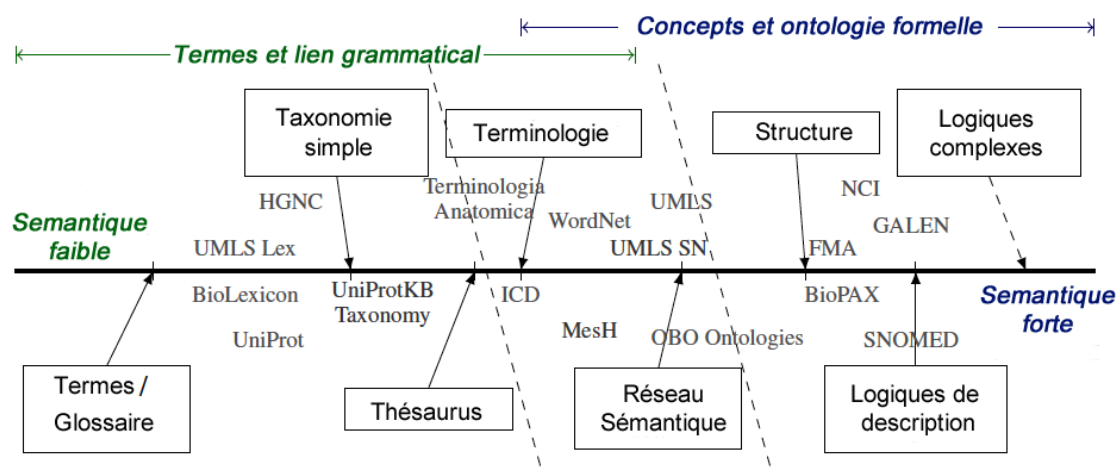


FIGURE 2.5 – Architecture de connaissance pour la construction d'une ontologie médicale

La Figure 2.5 classe les formalismes existants (Termes/Glossaire, Taxonomie simple, Thésaurus, ...) en fonction de leur expressivité sémantique. Les ressources biomédicales existantes sont placées en fonction du type de formalisme, par exemple les ressources lexicales (sur la partie gauche du schéma) contiennent la terminologie de plusieurs ressources avec des informations des concepts. Nous trouvons également des sources de lexiques spécialisés (par exemple UMLS) qui sont utilisées dans plusieurs applications de *text mining*. Les ressources plus complexes pour la construction des lexiques dans l'ontologie sont placées sur le réseau sémantique (par exemple NCI Metathesaurus, MeSH, SnomedCT, UMLS Metathesaurus, OBO ontologies, ...). Ce qui suit présente différents réseaux sémantiques existants.

UMLS¹⁵ (*Unified Medical Language System*) a été mis en place dans le but d'améliorer l'accès à l'information médicale à partir de diverses bases de données : bibliographiques, enregistrements cliniques et bases de connaissances médicales. UMLS sert à définir un

15. UMLS <http://www.nlm.nih.gov/research/umls/>

vocabulaire médical de base (méta-thésaurus) qui reprend et dédouble entre autres les termes de MeSH, SNOMED¹⁶. Ce méta-thésaurus propose une description hiérarchique des connaissances médicales utilisées dans divers documents et systèmes à base de connaissances.

Pour le traitement du vocabulaire médical en anglais, l'UMLS constitue un outil informatique puissant qui permet d'accéder à de nombreuses informations car il gère les variations des termes et les relations. Ce méta-thésaurus concerne essentiellement des terminologies anglophones. Il existe trois sources de connaissances UMLS : (1) un méta-thésaurus, qui contient des informations sémantiques sur des concepts biomédicaux, leurs noms et leurs relations ; (2) un réseau sémantique qui est un réseau de catégories générales ou de types sémantiques qui sont assignés à tous les concepts de méta-thésaurus ; (3) un lexique spécialisé qui contient des informations syntaxiques sur les termes biomédicaux et qui couvre la plupart des termes correspondant aux noms de concepts dans le méta-thésaurus.

FMA¹⁷ (*Fundational Model of Anatomy*) est une ontologie de référence pour le domaine de l'anatomie. C'est une représentation de toutes les entités anatomiques et les relations nécessaires pour la modélisation symbolique de la structure phénotypique du corps humain dans une forme qui soit compréhensible par l'homme et qui soit également computationnelle.

ICD¹⁸ (*International Statistical Classification of Diseases and Related Health Problems*) permet le codage des maladies, des traumatismes et de l'ensemble des motifs de recours aux services de santé. Il est publié par l'Organisation Mondiale de la Santé et est utilisé à travers le monde pour enregistrer (1) les causes et l'analyse systématique, (2) l'interprétation et la comparaison des données de mortalité et de morbidité recueillies dans différents pays ou régions à des époques différentes. La quantité de ces informations augmente chaque année.

MeSH¹⁹ (*Medical Subject Headings*) est un thésaurus biomédical de référence et est un outil d'indexation, de catalogage et d'interrogation des bases de données de la NLM (*National Library of Medicine, Bethesda, USA*), notamment MEDLINE/PubMed. Dans sa version 2012, la base de connaissance MeSH est constituée de 26 582 descripteurs avec 16 catégories thématiques, 150 descripteurs, 83 qualificatifs et environ 202 066 concepts supplémentaires. Il a été conçu à la *National Library of Medicine* aux États-Unis comme support de l'index Me, répertoire des principales publications scientifiques, et est utilisé par les systèmes de recherche bibliographique Medlars²⁰ et MEDLINE²¹. Le MeSH est organisé en deux parties : une liste alphabétique de termes (lexique) et une structure multi-axiale. Les termes équivalents sont rapportés à celui des 79 177 termes principaux (descripteurs) qui expriment le plus précisément le concept, termes auxquels

16. SNOMED <http://www.ihtsdo.org/snomed-ct/>

17. FMA <http://sig.biostr.washington.edu/projects/fm/>

18. ICD <http://www.who.int/classifications/icd/en/>

19. MeSH <http://www.ncbi.nlm.nih.gov/mesh>

20. Medlars <https://www.nlm.nih.gov/bsd/mmshome.html>

21. MEDLINE <https://www.nlm.nih.gov/bsd/pmresources.html>

sont associés un code alphanumérique. Les descripteurs s'organisent selon une structure hiérarchique et associative. Le MeSH comprend jusqu'à 11 niveaux de hiérarchie. Ces principaux composants sont les *Headings* (MH pour *Main Headings*), les *Subheading* et les *Supplementary Concept Records*. En outre, des connecteurs permettant des références explicites entre termes expriment les relations de synonymie, de voisinage ou d'association tandis que des qualificatifs permettent de considérer les différentes facettes d'un concept.

OBO²² (*Open Biomedical Ontologies*) est un travail collaboratif pour créer des vocabulaires contrôlés pour une utilisation partagée entre différents domaines biologiques et médicaux. En 2006, OBO faisait partie des ressources du *National Center for Biomedical Ontologies*. La bibliothèque d'ontologies OBO constitue la base d'une collaboration entre différents groupes de développeurs d'ontologies scientifiques qui ne cessent d'améliorer, d'augmenter et d'enrichir son contenu. Les ontologies présentes dans OBO ont été validées et sont utilisées par des experts du domaine. Elles sont considérées comme des ontologies formelles.

Le projet GALEN²³, développé à l'Université de Manchester, vise à mettre en place un serveur de terminologie en médecine. Développé depuis 1992 au sein de projets européens successifs, il est centré sur un modèle de références communes. GALEN utilise un formalisme appelé GRAIL²⁴ (*Galen Representation and Integration Language*) qui permet de saisir la connaissance terminologique dans le domaine médical. Ce formalisme est généraliste et permet de définir des concepts complexes (ou définis) composés de concepts plus élémentaires (ou primitifs). Tous les concepts, et les relations qui les lient, sont représentés indépendamment du langage dans lequel ils sont exprimés. GALEN est basé sur un modèle sémantique solide de terminologies cliniques appelé GALEN CORE²⁵ (*Coding Reference*). Ce modèle contient des concepts cliniques élémentaires (par exemple fractures, os, ...), des relations qui contrôlent la combinaison des concepts (par exemple les fractures osseuses) et des concepts complexes (par exemple fracture de la clavicule).

SNOMED²⁶ (*Systematized Nomenclature of Medicine*) est une nomenclature pluri-axiale couvrant tous les champs de la médecine et de l'odontologie humaine, ainsi que la médecine animale. Il s'agit d'un système de classification permettant de normaliser l'ensemble des termes médicaux utilisés par les praticiens de santé. La SNOMED a pour fonction d'attribuer un code à chaque concept permettant un grand nombre de combinaisons entre eux. Elle comprend également une liste de diagnostics au format numérique. Elle permet ainsi de stocker des informations médicales individuelles dans des entrepôts de données afin de concevoir des outils d'analyse décisionnelle, de faciliter des décisions thérapeutiques, de contribuer aux études épidémiologiques et à l'enseignement. L'utilisation de SNOMED garantit l'universalité du vocabulaire médical.

22. OBO <http://www.obofoundry.org>

23. GALEN <http://www.ncbi.nlm.nih.gov/pubmed/7889770>

24. GRAIL <http://www.ncbi.nlm.nih.gov/pubmed/9040895>

25. GALEN CORE <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233226/>

26. SNOMED <http://www.ihtsdo.org/snomed-ct/>

Kingsbury Center for Cancer Care Glossary a été obtenu par le *Kingsbury Information Center*²⁷. C'est un glossaire de termes liés au diagnostic et au traitement du cancer, spécialement destiné aux patients et à leurs familles. Ainsi *MedicineNet Medical Dictionary*²⁸ est un glossaire des termes médicaux pour les patients et les semi-spécialistes.

Multilingual Glossary of Technical and Popular Medical Terms in Nine European Languages²⁹ a été élaboré par l'*Heymans Institute of Pharmacology*. Ce glossaire comprend des termes médicaux en neuf langues de l'Union Européenne en ajoutant la dénomination correspondant en langage général (par exemple, antiseptique-désinfectant).

Il existe un nombre conséquent d'ontologies dans le domaine médical, qui facilite le processus de réutilisation d'ontologie. Malheureusement toutes les ontologies appliquées ou de domaines spécifiques (orthopédie, maladies vasculaires, ...) n'existent pas donc cela rend complexe l'étape de la réutilisation. Cela est dû à l'absence de stockage ou d'indexation unique. Nous pouvons citer les plus connues comme LOA³⁰ créée en 2008, ou encore comme OWG³¹, IO [Bir12], OBO, NCBO³². On trouve également des collections d'ontologies sur des pages personnelles partagées (comme Golbeck [Gol03]). On peut remarquer que le travail réalisé dans OBO vise à créer des ontologies biomédicales qui respectent le critère d'interopérabilité. Ces ontologies s'appuient sur un format standard : le format OBO, qui permet l'édition et la visualisation de celles-ci. Certaines ontologies restent cependant décrites au format OWL.

2.2.2/ DÉVELOPPEMENT DES ONTOLOGIES MÉDICALES

Le développement d'ontologies constitue un enjeu important. Les systèmes de traitement de l'information qui doivent fonctionner dans des domaines de connaissances spécialisés comme la médecine ne peuvent être efficaces que s'ils s'appuient sur des ressources terminologiques et ontologiques, construites pour le domaine concerné et en vue d'une application particulière [Bha09]. L'enjeu est d'élaborer des méthodes d'acquisition des connaissances à partir de textes qui spécifient : (1) comment utiliser les outils de traitement automatique des langues nécessaires pour l'analyse de corpus, et (2) les environnements de modélisation des connaissances nécessaires à la construction d'ontologies [Bha09, Pat09]. Quelle que soit la méthodologie adoptée pour le développement d'ontologie (cf. section 2.1.2), les processus de construction doivent faire l'objet d'une collaboration qui rassemble des experts du domaine à modéliser, des ingénieurs des connaissances et les futurs utilisateurs [Char12]. Cette collaboration ne peut être fructueuse que si les objectifs du processus de construction et les besoins qui en découlent sont clairement définis. L'ingénieur des connaissances doit s'interroger sur trois aspects du processus de construction : (1) l'objectif opérationnel, (2) le domaine de connaissance et (3) les utilisateurs.

La représentation conceptuelle des connaissances en médecine passe par le

27. Kingsbury Center <http://www.kingsburycenter.org>

28. MedicineNet, <http://www.medicinenet.com>

29. Multilingual Glossary, <http://glossarissimo.wordpress.com>

30. LOA, <http://www.loa.istc.cnr.it>

31. OWG, <http://mged.sourceforge.net/ontologies/>

32. NCBO, <http://www.bioontology.org>

développement d'ontologies. L'utilisation des ontologies dans la médecine se concentre principalement sur la représentation et l'organisation de terminologie médicale [Smi04]. Les médecins ont développé leur propre langage et lexique spécialisés pour les aider à stocker et à communiquer des connaissances médicales générales et des informations relatives aux patients de façon efficace. Ces terminologies sont optimisées pour la compréhension humaine et elles sont caractérisées par une quantité importante de connaissances implicites. Les systèmes d'information médicale doivent pouvoir communiquer des concepts médicaux complexes de façon détaillée (possiblement exprimés dans différentes langues, sans ambiguïté). C'est évidemment une tâche difficile et cela requiert une analyse approfondie de la structure et des concepts de la terminologie médicale.

D'après notre étude de la littérature, nous proposons de lister dix principes à respecter pour la construction d'ontologie :

1. **Domaine des connaissances** : il doit être délimité le plus précisément possible selon plusieurs axes ; les connaissances du domaine, les connaissances de raisonnement et les connaissances de haut niveau qui, par leur degré d'abstraction, peuvent être communes à plusieurs domaines [Cor06]. Le premier souci concernant la construction de l'ontologie est d'identifier quel est le domaine de connaissances que nous cherchons à modéliser (par exemple, si notre domaine de connaissance est le traitement du cancer, l'ensemble des connaissances que nous devons conceptualiser est l'analyse d'un corpus du domaine de la cancérologie). Quelle que soit sa spécialité médicale, un médecin se sert constamment de connaissances appartenant à d'autres spécialités. Dans ce cas, il est difficile de définir les limites des connaissances à modéliser et comment identifier les connaissances pertinentes.
2. **Degré de formalisme et granularité** : la formalisation est nécessaire pour permettre un raisonnement automatique afin de décharger les utilisateurs d'une partie de leur tâche d'exploitation [Char06]. Ces utilisateurs doivent, au même titre que l'objectif opérationnel, être identifiés car le choix du degré de formalisation et de la granularité de l'ontologie dépendent de leurs besoins et de leurs compétences. Les principales problématiques sur les questions de formalisme et de granularité concernent l'interprétation du sens des connaissances médicales en logique de description [Sch07, Bha09] :
 - comment passer de l'expression linguistique des connaissances à une représentation formelle et calculable, propre à l'exploitation informatique,
 - comment ne pas réduire de manière excessive l'expressivité du langage médical en le formalisant,
 - comment définir quelles sont les primitives de représentation et leurs significations dans le processus de modélisation.
3. **L'importation et la réutilisation** : le premier objectif est de définir la configuration de l'ontologie sur la base de discussions avec des experts du domaine et les sources de données existantes, comme les bases de données ou d'autres ontologies. En raison du nombre croissant de sources de connaissances dans le domaine biomédical, il est nécessaire de pouvoir mettre ensemble toutes ces sources complémentaires. L'automatisation de cette tâche est plus que nécessaire. La combinaison de différentes sources peut être triviale si un lien explicite entre les concepts similaires présents dans une base de données relationnelle existe. Selon la façon dont les ontologies sont combinées, deux principales techniques

sont définies par Noy [Noy00] : l'alignement et la fusion des ontologies. En alignement d'ontologies, les deux ontologies restent intactes et des liens entre elles sont créés. Cette technique est utilisée quand les ontologies sont de domaines complémentaires. Dans la fusion des ontologies, les deux ontologies sont fusionnées en une seule. Cette technique est utilisée quand il est nécessaire de prévoir une ontologie cohérente et unifiée.

4. **La classification hiérarchique** : elle est utilisée pour regrouper des termes en fonction du contexte. Les structures hiérarchiques résultantes peuvent être vérifiées par un ingénieur ontologique et/ou par un expert du domaine. Blaschke [Bla02] et Krishnan [Kri12] ont utilisé la classification hiérarchique pour grouper des termes du domaine provenant de la littérature. Le résultat a été un ensemble d'arbres indépendants qui ont été fusionnés par un expert du domaine. Ces arbres hiérarchiques qui représentent la taxonomie dans l'ontologie peuvent être conçus selon trois principes :

- l'identification des concepts les plus génériques que l'on déclinera en concepts de plus en plus spécifiques. Il s'agit d'une approche de haut en bas (ou TOP DOWN),
- l'identification des concepts spécifiques que l'on organise avec des concepts plus génériques. C'est une approche de bas en haut (ou BOTTOM UP),
- l'identification des concepts les plus importants (pas forcément spécifiques ou génériques) et à partir de ceux-ci l'identification des concepts plus génériques et plus spécifiques dont on aura besoin. Cette approche part du milieu vers les extrémités (ou MIDDLE OUT).

Dans la pratique, il n'y a pas d'approche purement « *TOP DOWN* » ou « *BOTTOM UP* » surtout lorsqu'une ontologie est conçue par une combinaison de constructions taxonomiques ou à partir d'ontologies déjà existantes réutilisées.

5. **Les ontologies représentent « ce qui est », les modèles d'information représentent « ce que nous savons »** : la tâche de représentation des faits cliniques exige deux aspects strictement séparés : (1) Les vérités universelles sur les entités du monde visé par des termes de domaine, et (2) les faits connus sur des cas cliniques concrets. La première tâche correspond à ce que l'on entend par ontologie, la seconde correspond aux modèles d'information. Il s'agit non seulement des faits qui sont dans le monde, mais aussi de la connaissance de ces faits. Par exemple il existe une distinction entre les classes *Tuberculose pulmonaire, confirmée par culture médical* et *Tuberculose pulmonaire, confirmée par antécédents cliniques*, or la tuberculose pulmonaire est confirmée comme maladie (vérité universelle), le patient X a d'un point de vue pulmonaire des facteurs à risque de tuberculose (connaissance). La nature de la tuberculose d'un patient ne dépend pas de la façon dont elle est diagnostiquée. Cette distinction n'a pas de sens au niveau des ontologies médicales car l'information ne change pas en fonction d'un fait connu ou inconnu [Rec06].
6. **Les exigences pratiques peuvent justifier les écarts contrôlés** : dans de nombreux cas, les besoins des utilisateurs peuvent engendrer des difficultés dans la conception des ontologies. Comme les ontologies sont basées sur des vérités universelles, elles évitent la représentation des conjectures qui sont importantes dans le contexte médical (par exemple, il peut être nécessaire de classer certains facteurs de risques liés aux maladies : si nous déclarons « *Hypertension* » comme « *Facteur de risque pour une rupture d'anévrisme* » dans une ontologie en utilisant

un relation « *is_a* », nous allons certainement la représenter comme une vérité universelle, car toute hypertension provoque un infarctus du myocarde. Ainsi confondre les rôles pour les relations de sous-classe est une erreur commune dans la conception de l'ontologie car l'hypertension n'est pas un symptôme obligatoire pour un accident vasculaire)[Sta09]. Pour cette raison, il est recommandé de concevoir l'ontologie avec des affirmations divergentes (Figure 2.6).

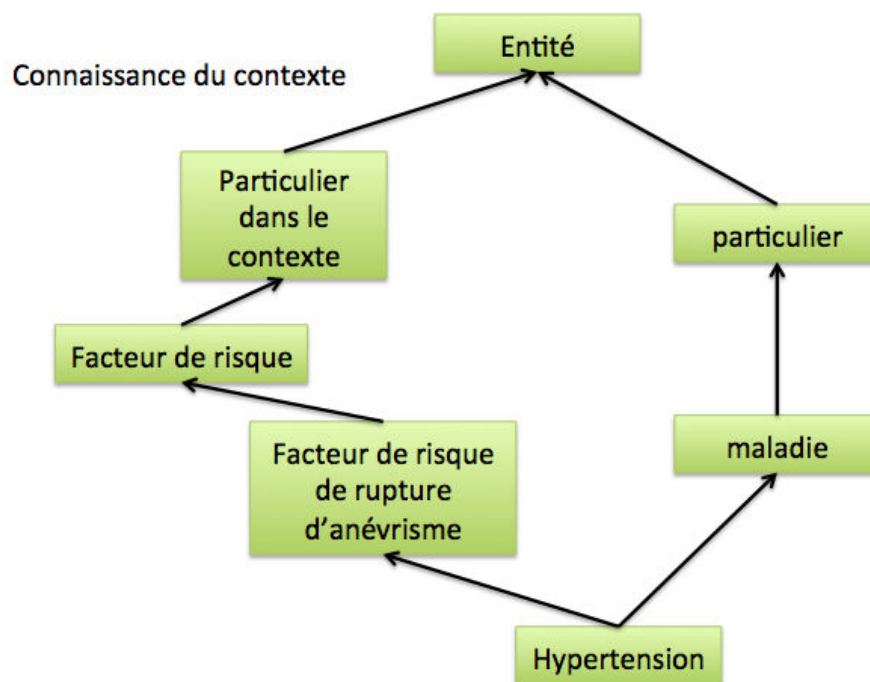


FIGURE 2.6 – Connaissances épistémologiques de l'ontologie

7. **Les ontologies doivent être liées à des dictionnaires** : lors de la création d'une ontologie, il n'est pas obligatoire d'intégrer un lexique ou des informations terminologiques même si cela est fortement recommandé. Cependant, il est important que les nœuds de l'ontologie soient eux-mêmes explicites et emploient des termes couramment utilisés dans le domaine. En outre, ils doivent éviter des formulations ambiguës. Cela signifie que les termes de l'ontologie ne doivent pas être erronés ou porter à confusion. Deux nœuds ne peuvent pas avoir le même terme. Il est alors parfois nécessaire de choisir des synonymes. C'est pourquoi il est nécessaire de fournir un lien entre l'ontologie et un dictionnaire où chaque entrée du dictionnaire correspond à un terme du domaine ou à des synonymes. Ce dictionnaire est une structure de données séparée et ne fait pas partie intégrante des ontologies.
8. **Le découpage d'ontologie** : les étapes du développement de l'ontologie présentées précédemment recueillent l'information et elles peuvent exiger le découpage du contenu de l'ontologie pour maintenir l'information qui correspond le mieux au domaine. Nous devons tenir compte de l'équilibre entre intégralité (qui peut poser des problèmes dans la gestion du contenu) et complexité du traitement (qui limite l'expressivité de l'ontologie). Lee [Lee11] propose le découpage sur l'exemple d'une ontologie développée sur le domaine des drogues. Il propose de

supprimer les concepts non nécessaires dans l'ontologie du domaine. Les concepts supprimés ne doivent pas changer la similarité sémantique sur l'ontologie complète.

9. **Les utilisateurs des ontologies n'ont pas besoin de connaître l'ensemble du système** : du point de vue de l'utilisateur, la structure interne d'une ontologie peut facilement conduire à la confusion ou à des interprétations erronées. Les utilisateurs inexpérimentés dans l'ingénierie ontologique n'ont pas besoin de connaître tous les détails de la hiérarchie. Ainsi, les ontologies ne doivent montrer aux utilisateurs que l'organisation des catégories bien définies pour faire respecter la cohérence dans le contenu. Il faut fournir des interfaces sur lesquelles s'appuyer pour des tâches personnalisées.
10. **Les ontologies doivent fournir des services de terminologie personnalisés** : les utilisateurs finaux des ontologies médicales doivent recevoir des terminologies précises, sans avoir besoin d'accéder à l'ontologie, car ces utilisateurs créent et sélectionnent uniquement des termes. Le service de terminologie est à son tour responsable de la liaison des termes de l'ontologie pour la représentation contrôlée à l'utilisateur (Figure 2.7).

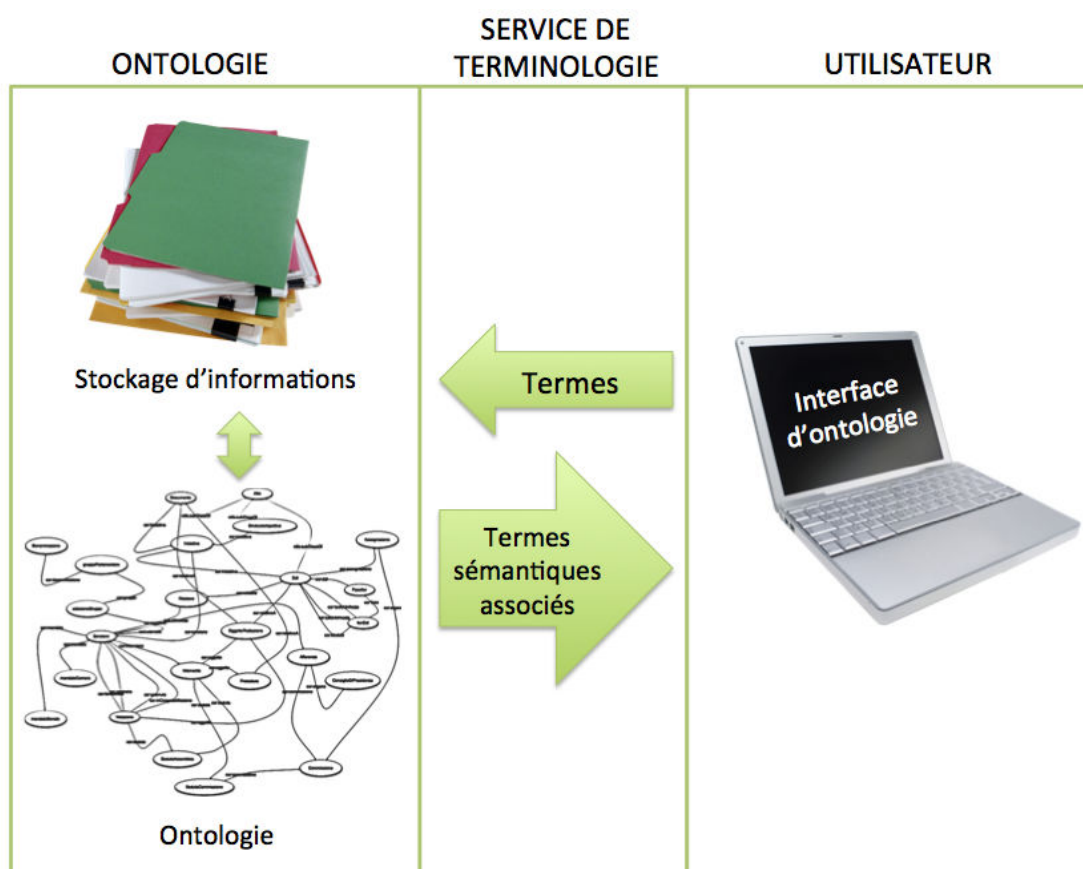


FIGURE 2.7 – Services de terminologie en médiation entre l'ontologie et l'utilisateur final

La représentation des connaissances dans le contexte clinique est étudiée par la communauté d'informatique médicale depuis la dernière décennie [Cim00]. L'objectif est d'offrir un ensemble de références sémantiques sur les connaissances cliniques (pathologies, traitements, médicaments, ...). Ces ensembles comptent actuellement différents

systèmes de classification de terminologies sur une large gamme de disciplines cliniques, grâce au fort enrichissement d'UMLS et à la croissance du développement d'ontologies en OBO. Pourtant, la plupart de ces terminologies et ontologies sont créées dans des contextes d'applications spécialisées. La plupart ne sont pas destinées à l'interopérabilité sémantique (elles couvrent tous les vocabulaires de contrôle des différentes disciplines de santé). Cette interopérabilité constitue une exigence incontournable dans le domaine de l'intégration des soins, où les données du patient doivent être facilement interchangeables à travers les frontières institutionnelles (cliniques, hôpitaux, centres de recherche, ...). L'objectif principal des ontologies biomédicales est de concentrer les efforts de représentation et de développement d'ontologies dans ce type de systèmes spécialisés sur la contribution médicale dans un contexte de soins cliniques intégrés.

Une ontologie médicale sur les maladies, les traitements, la coordination médicale, est initialement développée pour modéliser et représenter l'information issue des articles médicaux, des bibliothèques digitalisées, des diagnostics numérisés, ... Elle peut initialement être construite par taxonomies médicales et en réutilisant des ontologies existantes (UML, MeSH, NCI Thesaurus, ...). Les concepts supplémentaires et la sémantique relationnelle peuvent être enrichis avec l'exploration des ontologies existantes qui ont un lien avec le domaine d'intérêt, qui favorisent des détails supplémentaires qui peuvent avoir un intérêt pour le traitement, le dosage, les effets, les causes, les résultats de recherches. Il existe dans la littérature plusieurs méthodologies proposées pour le développement des ontologies (*cf.* section 2.2). La plupart d'entre elles ont été conçues pour résoudre des difficultés particulières dans le développement des ontologies à grande échelle (par exemple la gestion d'équipes coopératives en fonction des distances géographiques, offrant d'énormes sources d'acquisition de connaissances, et la fusion/alignement des ontologies hétérogènes) [Gom07]. Ces méthodes n'ont pas été adoptées pour la construction d'ontologies dans le domaine médical, malgré le fait qu'elles puissent être utilisées pour construire des ontologies de petite ou grande taille. La conception ontologique reste encore en partie un art quel que soit les méthodes utilisées [Fer07].

Les ontologies sont généralement développées sur un principe de saturation [Kaz11]. L'idée de base est de construire l'ontologie par un apprentissage au travers de documents cliniques. Ensuite, l'ensemble des informations est utilisé comme une phase de test et d'évaluation de l'ontologie qui couvre la plupart des concepts et des relations du domaine établi. Si de nouveaux concepts sont introduits durant la phase de test, les documents de tests établis doivent être mis à jour et les nouvelles classes et propriétés sont ajoutées à l'ontologie. Un deuxième test est utilisé pour évaluer l'ontologie de manière itérative. Ce processus a pour objectif de compléter les concepts, les propriétés et les relations. Souvent l'identification de nouveaux concepts (ou l'évolution de concepts existants) se fait manuellement sur la base des recommandations faites par les experts dans le domaine médical [Aba13]. Les concepts spécifiques souffrent parfois de généralisations dérivées des classes et des relations (différents procédés médicaux appliqués dans différents milieux), c'est l'identification des traitements, des médicaments, des symptômes, ... qui n'a pas été spécifiée dans l'ontologie. Il est important d'être consistant dans sa logique de développement.

Par exemple, une ontologie, qui modélise des maladies et/ou le traitement appliqué, doit être représentée par une série de classes ou de concepts qui sont directement liés à des propriétés ou des événements. Les classes hiérarchiques dans une ontologie doivent représenter la spécificité des relations (la représentation d'un traitement spécifique ou d'une maladie particulière dérivée de manifestations spécifiques, telles que des descrip-

tions d'articles médicaux). Les instances dans l'ontologie ne sont pas des traitements spécifiques pour la condition d'un patient en particulier, mais une information générique applicable à un groupe de patients avec une instance de maladie similaire et dans des conditions similaires. La Figure 2.8 montre un exemple de hiérarchisation générale pour la modélisation d'une pathologie et le traitement possible.

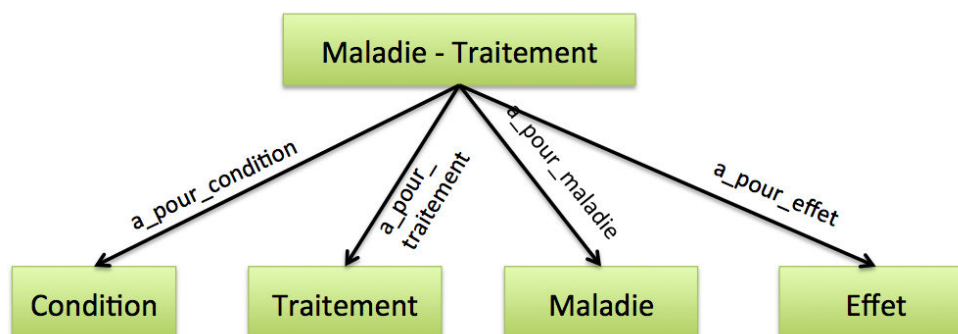


FIGURE 2.8 – Les classes de maladies/traitement dans une ontologie

Les Figures suivantes montrent une hiérarchie qui est considérée comme un complément d'un aspect médical, où « *maladie* » représente l'altération des fonctions de l'état physiologique ou morphologique considérées comme normales. Les maladies complexes sont dues à l'interaction entre un profil génétique particulier et un environnement (Figure 2.9). Le « *traitement* » suit l'évolution de la maladie concernée comme par exemple la prise de médicaments, les thérapies, les chirurgies (Figure 2.10). La « *condition* » se réfère à des conditions supplémentaires ou à des attributs que le patient pourrait présenter pour l'efficacité des traitements (Figure 2.11), ajoutant ainsi des données spécifiques aux patients (âge, sexe, les antécédents médicaux, ...).

Au sein des ontologies, si l'on utilise le langage OWL, il est possible d'associer aux propriétés (relations) des sous-propriétés (sous-relations) qui forment ainsi une hiérarchie de relations.

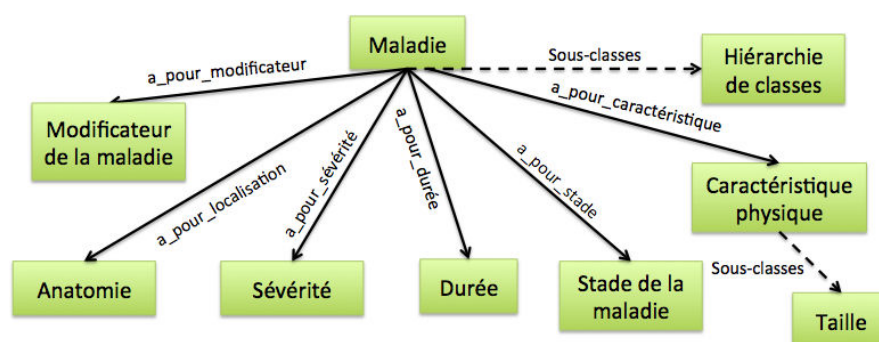


FIGURE 2.9 – Les classes de maladies dans une ontologie

Il est important de noter qu'il n'y a aucune limite sur le nombre de fois ou un composant peut être référencé. Par exemple, un traitement peut comprendre une combinaison de traitements (médicaments, chirurgie, ...), de la même façon, un traitement peut être appliqué à une combinaison de conditions pour un patient qui conduit à des effets multiples.

Une fois construite et acceptée par une communauté particulière, cette ontologie traduit

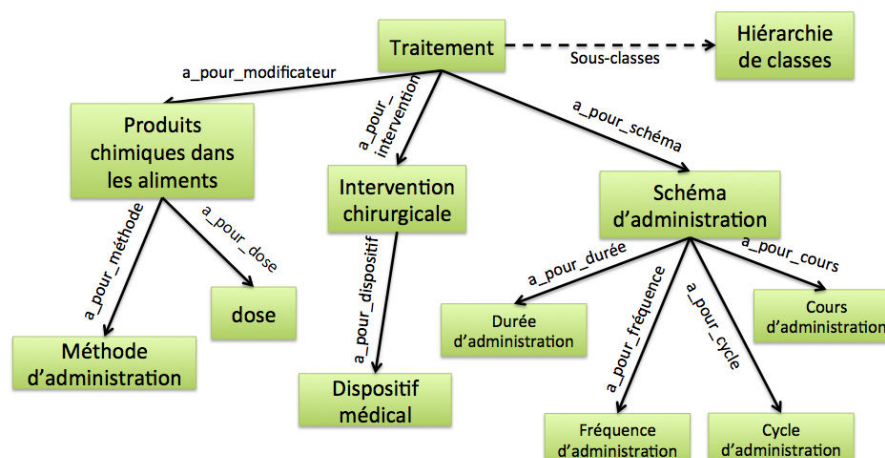


FIGURE 2.10 – Les classes de traitement dans une ontologie

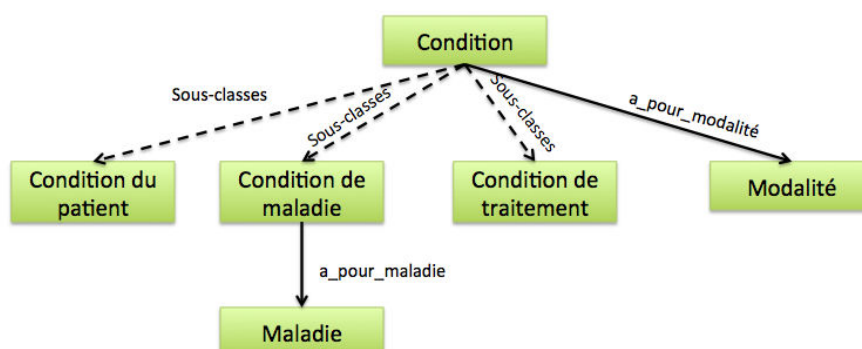


FIGURE 2.11 – Les classes de condition de maladie dans une ontologie

un consensus explicite et un certain niveau de partage : deux aspects essentiels pour permettre son exploitation par différentes applications ou agents logiciels. Ce point de vue soulève un paradoxe important : dans un souci de réutilisation, l'ontologie gagne à cultiver une certaine indépendance par rapport aux différentes applications dans lesquelles elle peut être utilisée et sa construction elle-même doit être guidée par l'usage dans l'application cible. Il se pose alors le problème de l'utilisation opérationnelle des ontologies, c'est à dire de leur mise en œuvre pratique [Pat09]. Il est indispensable de préciser l'objectif opérationnel de l'ontologie, en particulier au travers de scénari d'usage. Ainsi, l'étape, indispensable, de validation de l'ontologie doit être assurée par des professionnels de la santé.

Le contenu, la forme, la couverture et le degré de formalisation doivent être adaptés en fonction du rôle que doit jouer l'ontologie dans l'application cible. Une fois construite et acceptée par une communauté particulière, cette ontologie traduit un consensus explicite et un certain niveau de partage, deux aspects essentiels qui permettent son exploitation par différentes applications ou agents logiciels [Cim00].

2.2.3/ ONTOLOGIES APPLIQUÉES

Il existe différents types d'ontologies. Dans la littérature Guarino[Gua98] et Gómez-Pérez [Gom07] présentent plusieurs typologies connues qui classifient les ontologies en fonction : des objets qu'elles modélisent, de la granularité des connaissances représentées et du niveau de formalisme du modèle (Figure 2.12).

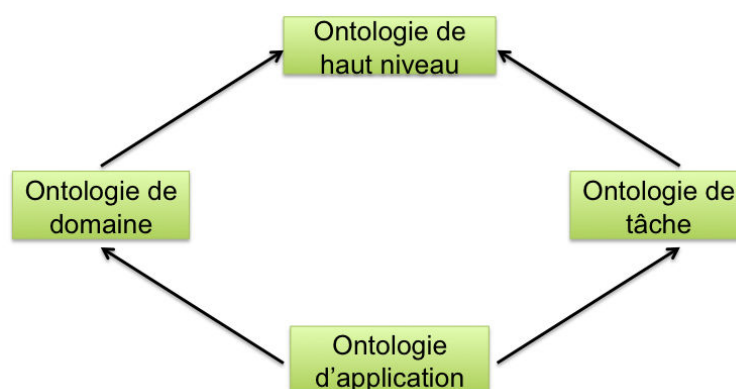


FIGURE 2.12 – Classification des ontologies selon Guarino [Gua98]

Ce type d'ontologie regroupe les concepts utilisés pour formaliser les connaissances. L'ontologie définit les concepts utilisés principalement dans des langages structurés (classes, sous-classes, attributs, valeurs, relations et axiomes) [Gru93a]. L'ontologie de domaine permet de spécifier les connaissances d'un domaine de façon indépendante du type de manipulation qui va être opéré.

De nombreuses ontologies de domaine existent déjà (Figure 2.13). MENELAS [Zwe99] présente des ontologies de haut niveau de conceptualisation. Le projet LERUDI (*LEc-ture Rapide en Urgence du Dossier patient Informatisé*)³³ vise à développer un système d'information offrant aux professionnels de la santé une vision synthétique du dossier patient informatisé et la possibilité de prises de décisions médicales. Le développement d'ONTOURGENCES [Char12] a été réalisé en 6 étapes :

- construction du squelette ontologique de la ressource termino-ontologique,
- utilisation de ressources terminologiques et ontologiques existantes pour compléter manuellement le système de concepts,
- enrichissement automatique,
- acquisition de concepts ou niveau de termes,
- enrichissement des termes en concepts en rapport avec les médicaments,
- mise en œuvre de procédures de validation.

Ces 6 étapes sont les règles de construction de toutes ces ontologies : les deux premières étapes de construction correspondant à une méthode classique de construction d'ontologie, les trois étapes suivantes étant en revanche spécifiques au développement de ressources termino-ontologique et à son usage dans un système. La validation est l'assurance qualité qu'il faut prendre en considération pour tous les types d'ontologies.

Dans la littérature, nous trouvons différentes ontologies appliquées dans différents domaines.

33. LERUDI, <http://esante.gouv.fr/actus/services/le-projet-lerudi-fiche-signalétique>

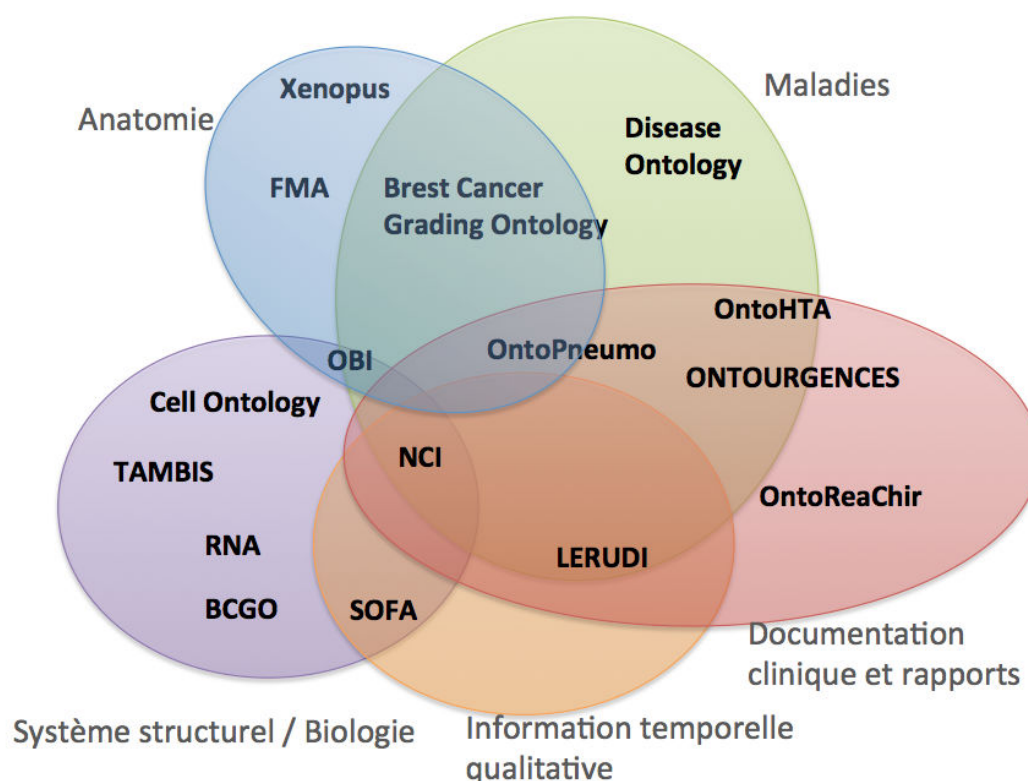


FIGURE 2.13 – Ontologies existantes dans le domaine médical

Pour la conception de systèmes biologiques : l'ontologie de TAMBIS (Tao) est une ontologie pour les concepts de la biologie moléculaire et de la bio-informatique. Elle a été créée comme une partie d'un système complet de requêtes qui a le même nom (TAMBIS). Elle a été modélisée avec le langage *DAML+OIL*. Elle est utilisée pour décrire : (1) les méta-données de sources de données, ce qui représente un schéma universel sur les fichiers, (2) la formation des requêtes sur le langage de modélisation qu'ils ont créé, (3) l'interface utilisateur pour la formulation de requêtes, elle est destinée aux utilisateurs qui ne sont pas préparés pour utiliser les langages de requêtes complexes, (4) la médiation entre différentes sources pour traduire le modèle médiateur en modèle source. Une caractéristique importante de cette ontologie est qu'elle ne contient pas d'instances, c'est à dire qu'elle ne contient que des connaissances sur des concepts bio-informatiques et de biologie moléculaire ainsi que leurs relations. Il existe également une ontologie du cycle cellulaire pour représenter le processus et l'analyse des composants de réseaux moléculaires liés au cycle cellulaire [Ant09]. De façon similaire, il existe une ontologie de composants cellulaires pour aider la modélisation des propriétés et des composants fonctionnels des cellules [Mat10].

Pour la gestion de données et de l'information : malgré ses débuts comme méta-thésaurus basé sur UMLS, les gestionnaires de NCI (*National Cancer Institute*)³⁴ ont décidé de le transformer en une ontologie [Gon11]. NCI contient la terminologie produite par l'Institut National sur le Cancer. Il est l'un des plus grands sujets décrivant les maladies, les médicaments, les produits chimiques, les diagnostics, les gènes, les thérapies,

34. NCI, <http://www.cancer.gov/cancertopics/cancerlibrary/terminologyresources>

l'anatomie, les organismes et les protéiques. De façon similaire, une ontologie du cancer de la prostate a été conçue à l'aide de la gestion et de l'intégration des données cliniques sur les diagnostics de maladies cancéreuses [Min09]. Il existe également une ontologie sur la gestion des concepts liés aux traitements et à la recherche du cancer, pour la gestion et l'intégration des données des essais cliniques [Gei13].

Pour le support des maladies spécifiques et de la prise de décision clinique : l'ontologie des maladies (*Disease Ontology*)³⁵ correspond à un vocabulaire médical contrôlé développé dans le champ bio-informatique en collaboration avec le projet NuGene du Centre de Médecine Génétique, de l'Université de Northwestern, Chicago. Elle a été conçue pour fournir une trace entre les maladies et les conditions associées à certains codes médicaux comme entre autres ICD9CM³⁶ et SNOMED³⁷. L'ontologie OntoPneumo a été développée pour aider au codage médical dans le domaine de la pneumologie [Ban08]. La construction de cette ontologie a fortement utilisé les ressources terminologiques, non seulement des pratiques médicales en pneumologie mais également des vocabulaires utilisés par les médecins (l'ontologie a été validée par un expert du domaine, médecin hospitalier dans un service de pneumologie). L'ontologie OntoHTA permet pour la mise à jour des formulaires pour des données cliniques dans le domaine de l'hypertension artérielle [Ste07]. OntoReaChir a été développée pour la prise en charge des complications postopératoires et traumatologie dans le domaine de la réanimation chirurgicale [LeM02]. Ces ontologies médicales françaises sont des hiérarchies strictement taxonomiques, dans le respect des principes de la sémantique différentielle. Elles sont également construites avec la formalisation OWL et bénéficient d'une *top level* ontologie (une ontologie formelle ayant un haut niveau d'abstraction) articulant l'ensemble des concepts.

Pour le support des recherches biologiques : l'ontologie pour la recherche biomédicale (OBI)³⁸ a pour principal objectif le développement d'une ontologie intégrée pour la description des recherches biologiques et cliniques. Elle modélise la conception d'une recherche sur les protocoles et les instruments à utiliser, le matériel utilisé, les données générées et le type d'analyses effectuées. Elle est considérée comme une ontologie de OBO³⁹.

Pour le développement d'ontologies basées sur la réutilisation : l'ontologie de RNA⁴⁰ (*RiboNucleic Acid ou Acide ribonucléique - ARN*) a pour finalité de créer une ontologie pour comprendre le fonctionnement des séquences d'ARN dans des systèmes biologiques. Un des objectifs de ce projet était de créer une ontologie en s'appuyant sur trois ontologies existantes : l'ontologie Gene (*GO*), l'ontologie de séquence (*SO*) et l'ontologie structurelle utilisée par PDB (*Protein Data Bank*). *GO* fournit des vocabulaires contrôlés qui décrivent les gènes et la qualité des produits génériques dans n'importe quel organisme. Celle-ci est également composée de trois ontologies : une qui décrit des

35. Disease Ontology, <http://diseaseontology.sourceforge.net>

36. ICD9CM <http://www.cdc.gov/nchs/icd/icd9cm.htm>

37. SNOMED <http://www.ihtsdo.org/snomed-ct/>

38. OBI, <http://obi.sourceforge.net/index.php>

39. OBO <http://www.obofoundry.org>

40. RNA, <http://roc.bgsu.edu>

produits génétiques comme leur processus biologique associé, une seconde pour les composants cellulaires, et la dernière pour décrire les fonctions moléculaires de façon indépendante de l'espèce. Un autre aspect important de GO est d'utiliser un cycle de construction simple et intuitive. SO et PDB cherchent à développer une ontologie qui convient pour décrire des séquences biologiques. Il existe une version allégée de cette ontologie appelée SOFA (*Sequence Ontology Feature Annotation*)⁴¹. Cette ontologie est écrite dans le langage OWL.

Il est important de noter la grande quantité d'information existante sur les ontologies dans le domaine médical, ce qui peut être considéré comme un avantage. Il est possible d'obtenir une grande compréhension de la terminologie médicale en utilisant efficacement ces nombreux systèmes qui s'appuient sur ces ontologies à des fins différentes. Notons également qu'OWL est le langage pionnier de modélisation de l'ontologie.

2.3/ CYCLE DE VIE D'UNE ONTOLOGIE

Il s'agit dans cette section d'étudier le cycle de vie des ontologies, autrement dit comment ont évolué les ontologies médicales, comment elles peuvent évoluer et enfin comment elles sont maintenues.

Puisque les ontologies sont destinées à être utilisées comme des composants logiciels dans des systèmes répondant à des objectifs opérationnels différents, leur développement doit s'appuyer sur les mêmes principes que ceux appliqués en génie logiciel. Ainsi, les ontologies doivent être considérées comme des objets techniques évolutifs possédant un cycle de vie qui nécessite d'être précisé. Les étapes pour une ontologie comprennent des activités de gestion de projet (planification, contrôle, assurance qualité) et des activités de développement (spécification, conceptualisation, formalisation) auxquelles s'ajoutent des activités transversales de support telles que l'évaluation, la documentation et la gestion de la configuration [Gan06].

Le cycle de vie d'une ontologie (Figure 2.14) comprend :

- une étape initiale de détection et de spécification des besoins qui permet notamment de circonscrire précisément le domaine des connaissances,
- une étape de conception découpée en trois sous-étapes : Normalisation, Formalisation et Opérationnalisation,
- une étape de déploiement et de diffusion,
- une étape d'utilisation incontournable,
- une étape d'évaluation,
- et une étape d'évolution et de maintenance du modèle.

L'ontologie et les besoins des utilisateurs doivent être réévalués et l'ontologie peut être ou bien étendue ou bien si nécessaire en partie reconstruite. La validation du modèle de connaissances est au centre du processus et se fait de manière itérative [Fer00].

Les phases de conception initiale et d'évaluation ont elles aussi un certain nombre de points incontournables :

- la spécification des solutions,

41. SOFA, <http://www.sequenceontology.org>

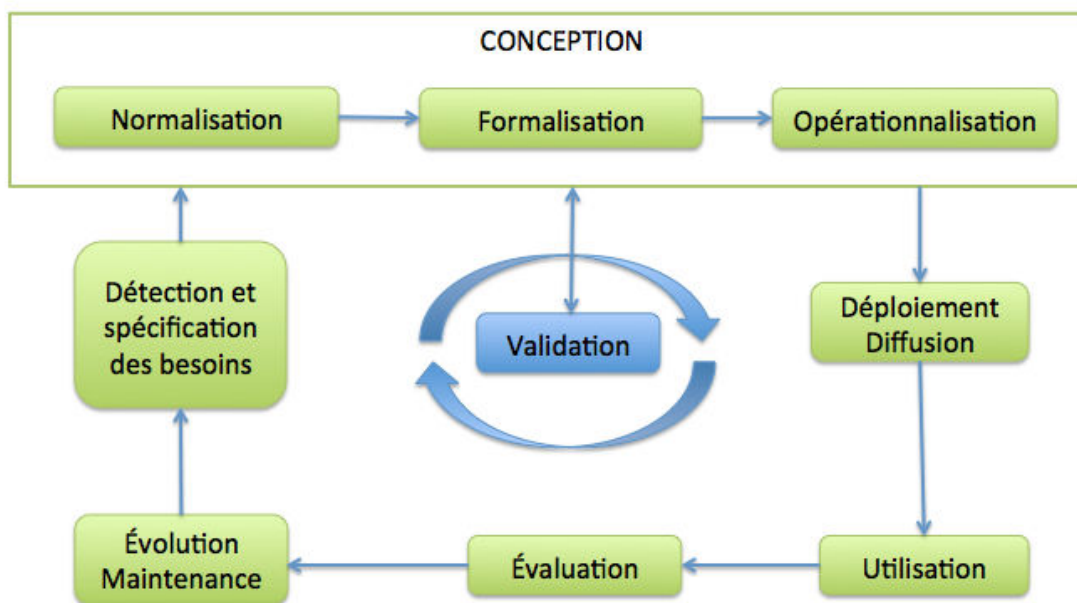


FIGURE 2.14 – Cycle de vie d'une ontologie

- l'acquisition des connaissances nécessaires (analyse de textes, traitement automatique de la langue naturelle, plateformes collaboratives),
- la conceptualisation et la modélisation (*design pattern* ontologiques, méta-ontologies, entretien avec les experts),
- la formalisation (méthodes et outils de l'ontologie formelle, logiques de description et algorithmes de tableaux, analyse formelle de concepts, graphes conceptuels, formalismes du web sémantique RDF et OWL),
- l'intégration de ressources existantes (alignement automatique d'ontologies, traduction),
- l'implantation (graphes conceptuels, logiques de description, formalismes objets).

Un autre problème lié à la conception et à l'évolution d'une ontologie est l'obtention et le maintien d'un consensus sur les choix de représentation et de conceptualisation faits dans l'ontologie. Suivant les usages, ce problème fait appel à des collecticiels (groupware, cf. chapitre 1) et des outils de gestion des points de vue, des terminologies, des langues et des jargons différents. Notons aussi que l'évolution pose le problème de la maintenance de l'ontologie existante. Lorsque l'ontologie est modifiée, ses changements ont un impact sur tout ce qui a été construit précédemment. Le maintien de la cohérence dans une ontologie, l'historique et la gestion des versions, la ré-ingénierie et la propagation des changements après modification, sont des questions de recherche encore largement ouvertes. La maintenance de l'ontologie soulève donc des problèmes d'intégration technique et des problèmes d'intégration aux usages [Gang06].

La médecine étant en constante évolution, la modélisation des connaissances et des pratiques dans ce domaine se situe entre invariabilité et évolution, entre stabilité et adaptabilité [Del13]. Les réflexions menées autour des questions de l'évolution, de la maintenance et des critères spécifiques à l'évolution de ressources de type ontologique apparaissent dans les communautés de recherche liées à la terminologie et l'ingénierie des connaissances.

L'évolution d'une ontologie peut avoir plusieurs objectifs : comme effectuer un bilan d'ensemble, ou guider la prochaine étape de construction, ou découvrir l'origine des difficultés, ou rendre compte de résultats . . . Il est clair que ces objectifs peuvent s'opposer et qu'ils doivent être organisés. Il faut évaluer par une méthode qualitative et quantitative la hiérarchisation relationnelle incluant les concepts et la taxonomie. Il faut également choisir entre évaluer toutes les étapes de construction ou bien uniquement le résultat obtenu.

Les critères pertinents pour évaluer une ressource terminologique et ontologique sont ceux qui doivent être respectés lors de la construction de cette même ressource. Ainsi, il est important de respecter un certain nombre de contraintes pour obtenir une modélisation de qualité, à la fois adaptable et maintenable [Rou10] :

- clarté et objectif : l'ontologie doit fournir la signification des termes définis en donnant des définitions objectives. Quand une définition peut être axiomatisée, elle doit l'être. Des définitions en langage naturel doivent être fournies,
- perfection : une définition exprimée par des conditions nécessaires et suffisantes est préférable à une définition partielle,
- cohérence et extensibilité : une ontologie doit être cohérente pour permettre des inférences conformes aux définitions. Les axiomes doivent être consistants. La cohérence des définitions en langage naturel doit être vérifiée. L'ajout de nouveaux concepts à l'ontologie ne doit pas entraîner la révision des définitions existantes. Plus généralement, l'ontologie doit être construite de telle manière que l'on puisse l'étendre facilement, sans remettre en cause ce qui a déjà été fait,
- encodage minimal : l'ontologie doit être conceptualisée indépendamment de tout langage d'implémentation. Le but est de permettre le partage des connaissances contenues dans l'ontologie entre différentes applications utilisant des langages de différentes représentations,
- ontologies minimales : une ontologie doit faire un minimum d'hypothèses sur le monde. Elle doit contenir un vocabulaire partagé mais ne doit pas être une base de connaissances comportant des connaissances supplémentaires sur le monde à modéliser. Elle doit choisir de représenter un point de vue en particulier sur le domaine qui l'intéresse.
- distinction ontologique : l'ontologie doit identifier et isoler un noyau de propriétés considérées comme invariables pour une classe (critère d'identité), il faut créer une nouvelle classe de concepts,
- normaliser l'ontologie : l'ontologie doit être normalisée chaque fois que cela est possible, ces normalisations représentant les classes des objets reconnus comme existant dans le domaine. Construire une ontologie, c'est aussi décider d'une manière d'être et d'exister des objets.

2.4/ ÉVALUATION D'UNE ONTOLOGIE

Comme tout logiciel, le contenu des ontologies (concepts, taxonomie et axiomes) doit être évalué avant d'être utilisé ou réutilisé dans d'autres ontologies ou dans des applications. Dans [Gom07], l'auteur indique que l'évaluation d'une ontologie est un processus technique de son contenu par rapport à un cadre de référence (les exigences, des questions de compétence, . . .), pendant toutes les phases de son cycle de vie (le bon usage du langage utilisé pour la codification, l'exactitude de la structure taxonomique, le sens des termes, les concepts représentés et l'adéquation des exigences exprimées au début

du développement). Ce processus inclut la vérification et la validation de l'ontologie et doit être effectué sur les définitions et les axiomes établis et qui peuvent être contrôlés à partir d'autres applications.

La vérification se réfère à la construction correcte d'ontologie, dit autrement, il faut s'assurer que les définitions implémentent correctement les exigences de l'ontologie. Pendant cette partie, la validation vise à assurer que les définitions ontologiques modélisent le monde réel pour lequel l'ontologie a été créée.

Durant le processus d'évaluation, une série de tests est établie et exécutée. Les résultats sont ensuite analysés [Tar10] à partir de trois états possibles des ontologies :

- pré-modélisation : cet état est considéré comme la révision et l'évaluation des matériels disponibles par le développement,
- modélisation : cet état vérifie la qualité des significations, de la cohérence et de la redondance des concepts, en utilisant des ontologies et d'autres questions de compétence. Il évalue également la possibilité d'erreurs syntaxiques faites pendant le codage,
- et libération : cet état détermine la qualité de l'ontologie en comparaison à d'autres ontologies différentes mais équivalentes.

Il existe de nombreuses propositions d'évaluation d'ontologie, telles que celles proposées dans [Bre04, Por04, Obr07, Weic09]. Ils proposent d'examiner les ontologies en respectant trois niveaux de base : niveau vocabulaire, niveau taxonomie, et niveau non-taxonomique basé sur des relations sémantiques. Ces niveaux sont également soumis à différentes approches d'apprentissage. L'évaluation d'ontologies de domaine, par exemple dans le domaine médical, repose sur une évaluation des quatre critères suivants : la rigueur taxonomique, le langage utilisé pour le codage, l'utilité d'application ou de tâche qui est utilisé pour l'ontologie, et le vocabulaire utilisé pour représenter les concepts et les relations du domaine modélisé [San12]. Dans [Sar11] les auteurs remarquent que la plupart de ces critères sont évalués sur la base d'une comparaison avec d'autres ontologies disponibles, qui sont utilisées comme une référence ou un standard. Cependant, lorsque l'ontologie en cours de développement est la seule description connue du domaine, les développeurs doivent utiliser des méthodes disponibles sans garantie que l'évaluation soit suffisamment fiable et complète.

Il est possible d'évaluer la qualité d'une ontologie sans l'obligation de références précédentes en s'appuyant sur l'examen d'un ensemble minimum de critères tels que : le vocabulaire utilisé pour représenter les connaissances a une couverture suffisante du corpus, l'ontologie est décrite d'une manière orthographiquement correcte (sans erreur et en conformité avec les règles de la langue utilisée), la structure taxonomique qui organise les concepts et les termes du domaine sont corrects, consistants et sans redondances, les conditions pour lesquelles l'ontologie a été créée sont satisfaites, et les questions de compétence ont une réponse correcte (par exemple, quel est le domaine que va couvrir l'ontologie, dans quel but utiliserons-nous l'ontologie, qui va utiliser et maintenir l'ontologie, ...) [San12].

Le développement et l'adaptation d'ontologies dans le domaine médical représentent une tâche complexe qui exige un effort considérable et une collaboration constante entre les experts (professionnels de la santé) et les ingénieurs ontologiques (experts de développement sémantique). Si les outils et les techniques de développement d'ontologies sont suffisants pour supporter le travail de conception, ils n'offrent que des solutions partielles pour la représentation des connaissances [Aba13]. Les ontologies, pour sou-

tenir la prise de décisions cliniques, doivent nécessairement inclure un soutien médical qui vérifie la représentation des activités dans la recherche de la qualité des effets et les réponses ontologiques sur le domaine à traiter. En conséquence, le développement automatique des ontologies médicales (les schémas de connaissances et la description des individus) doit être validé par les experts du domaine [Rui11].

Les experts médicaux ne sont pas familiarisés avec les formalismes et la technologie de l'ingénierie ontologique. Il est donc fortement recommandé de faire appel à des ingénieurs pendant le processus de validation, afin de réduire les erreurs potentielles. La validation des aspects logiques et structurels de l'ontologie comme l'incohérence, l'incomplétude ou la redondance, peut être diminuée automatiquement avec des outils spécialisés. Nous pouvons citer ici les trois plus connus :

- *Protégé-2000*⁴² est le plus connu et le plus utilisé des éditeurs d'ontologie, il est développé par l'Université Stanford et est *Open-source*,
- *SWOOP*⁴³ est un éditeur d'ontologie développé par l'Université du Maryland,
- *KMgen*⁴⁴ est un éditeur d'ontologie pour le langage KM (The Knowledge Machine). . .) [Vor10, Hua08].

CONCLUSION



Les ontologies sont des spécifications formelles qui permettent une représentation de la connaissance d'un domaine spécifique. Elles facilitent la communication, la réutilisation et le partage d'information entre les individus, les organisations et les systèmes informatiques. Les ontologies peuvent être considérées comme des ressources fondamentales pour le web sémantique. Pour cette raison, au cours des dernières décennies, les travaux de recherche se sont multipliés et ce en particulier dans le domaine médical.

Ce chapitre a présenté les aspects fondamentaux des ontologies générales et biomédicales. Nous avons examiné certaines méthodologies pour le développement des ontologies, comme Methontology et Ontology development 101 (méthodologies que nous avons utilisées pour nos travaux). Nous avons également présenté les langages ontologiques, le raisonnement des ontologies dans le domaine médical, les principes de conception ainsi que le cycle de vie et l'évaluation d'une ontologie.

42. Protégé 2000, <http://protege.stanford.edu>

43. SWOOP, <https://code.google.com/p/swoop/>

44. KMgen, <http://www.algo.be/dev-logiciels.htmkmed>

CONCLUSION DE LA PARTIE I



Dans cette première partie, nous avons présenté deux états de l'art.

L'état de l'art des télé-applications collaboratives a permis de situer le contexte de cette Thèse : celui de la télé-médecine et plus particulièrement du télé-diagnostic médical. Et dans ce contexte, de nouvelles exigences, en particulier du CNOM *Conseil National de l'Ordre des Médecins*, imposent la traçabilité des applications de télé-médecine : d'un point de vue médico-légal, il est nécessaire de pouvoir *rejouer* un diagnostic en sachant par exemple « qui a fait quoi et à quel moment ».

Comme les flux de données pris en compte dans un télé-diagnostic sont de taille très importante, il est apparu indispensable de faire appel aux ontologies pour développer de nouveaux outils de traçabilité qui seront capables de trouver de manière efficace et fiable, dans la masse d'informations, les informations pertinentes pour tracer le télé-diagnostic.

Ainsi, le chapitre suivant d'état de l'art a permis d'étudier les méthodes de conception, les outils et les langages qui permettent de construire des ontologies. Nous avons utilisé ces méthodes pour élaborer les trois ontologies qui constituent le cœur de notre plateforme *COOVADIS COllabOrative VAscular DiagnoSis*.

La deuxième partie de ce document est donc consacrée à la définition des trois ontologies, puis à celle de la plateforme complète. Nous présentons ensuite les validations (théorique et clinique) de *COOVADIS*.



CONTRIBUTION

INTRODUCTION

La deuxième partie de ce document de Thèse constitue notre contribution.

Dans un premier chapitre nous présentons les trois ontologies essentielles pour développer la traçabilité des diagnostics dans le domaine cardiovasculaire :

- ontologie du système vasculaire : ontologie que l'on peut qualifier de *classique* dans le sens où d'autres ontologies de pathologies existent (réalisée en collaboration avec les médecins de l'hôpital de Morelia au Mexique),
- ontologie de diagnostic et ontologie de traçabilité : qui sont deux ontologies très originales (que nous avons élaborées dans le cadre de nos travaux et qui sont l'une des contributions principales de cette Thèse). De plus, l'ontologie de traçabilité est générique dans le sens où le modèle pourra être reproduit dans le cadre d'autres pathologies (il suffira de développer une nouvelle ontologie pour une autre pathologie). L'ontologie de diagnostic, quant à elle, demeure encore liée à l'ontologie du système vasculaire.

Puis fondée sur ces ontologies, nous définissons dans le chapitre 4 notre nouvelle plateforme *COOVADIS COllabOrative VAscular DiagnoSis*. Développée en mode *SaaS Software as a Service*, elle permet de suivre et de tracer les diagnostics élaborés par un staff de médecins sur un patient atteint d'une pathologie cardiovasculaire, offrant ainsi la traçabilité des actes.

Enfin le dernier chapitre de cette partie expose les travaux de validation théorique, mais également de validation avec essai en milieu clinique, ainsi que les premiers résultats expérimentaux obtenus à l'aide de la plateforme *COOVADIS*.

CONCEPTION DES ONTOLOGIES DANS LE DOMAINE VASCULAIRE

INTRODUCTION

Les maladies vasculaires représentent la principale cause de décès dans le monde. Ces maladies sont diagnostiquées à l'aide d'une série d'analyses et d'imageries exploratoires. L'utilisation de l'imagerie médicale a permis aux médecins d'en apprendre plus sur le fonctionnement du corps humain. Ces images médicales suivent le standard DICOM (*Digital Imaging and Communication in Medicine*). Elles permettent également aux médecins d'intégrer leur diagnostic directement dans les images. Ce type d'image contient des informations supplémentaires : l'identification du patient, les procédures de l'étude, la date de l'étude, l'identification du médecin, l'identification de l'institution ainsi que le diagnostic. Ces détails sont importants pour assurer l'intégrité de l'image, la bonne attribution du patient à l'étude, l'ordonnance des médecins et les dates des procédures.

Dans ce chapitre nous expliquons le développement des ontologies pour le traitement et le diagnostic des maladies vasculaires à partir de ressources médicales et d'imageries. Nous appliquons sur ces ressources des techniques appartenant au domaine du traitement du langage pour construire l'ontologie de la pathologie. Nous construisons également des ontologies pour la traçabilité de la maladie du patient qui impliquent un diagnostic et une organisation temporelle du travail collaboratif entre médecins. L'objectif principal de l'aspect temporel est d'intégrer diverses données informatisées issues des dossiers de santé électroniques ainsi que l'imagerie médicale pour mettre en exergue l'évolution du patient.

Après avoir présenté, dans les deux premières sections, le contexte des maladies vasculaires, ainsi que les méthodes de conception de nos ontologies, nous avons organisé les deux dernières sections de la manière suivante : tout d'abord les deux ontologies liées au domaine vasculaire (ontologie du système vasculaire, et ontologie du diagnostic), puis l'ontologie de traçabilité.

3.1/ LES MALADIES VASCULAIRES

Comme nos travaux s'appliqueront au domaine cardio-vasculaire, il est nécessaire dans cette première section de donner au lecteur un minimum de connaissances sur cette pathologie.

Pour l'organisation mondiale de la santé (OMS¹), les maladies vasculaires sont la première cause de mortalité dans le monde. La prévalence des pathologies vasculaires (cardiaque, cérébral, pulmonaire, périphérique et rénale) est principalement causée par le durcissement des artères (athérosclérose) en raison d'un épaissement des parois artérielles par des dépôts graisseux, des coagulum de sang, de la calcification (athérome) ou également d'un rétrécissement des vaisseaux sanguins (Figure 3.1).

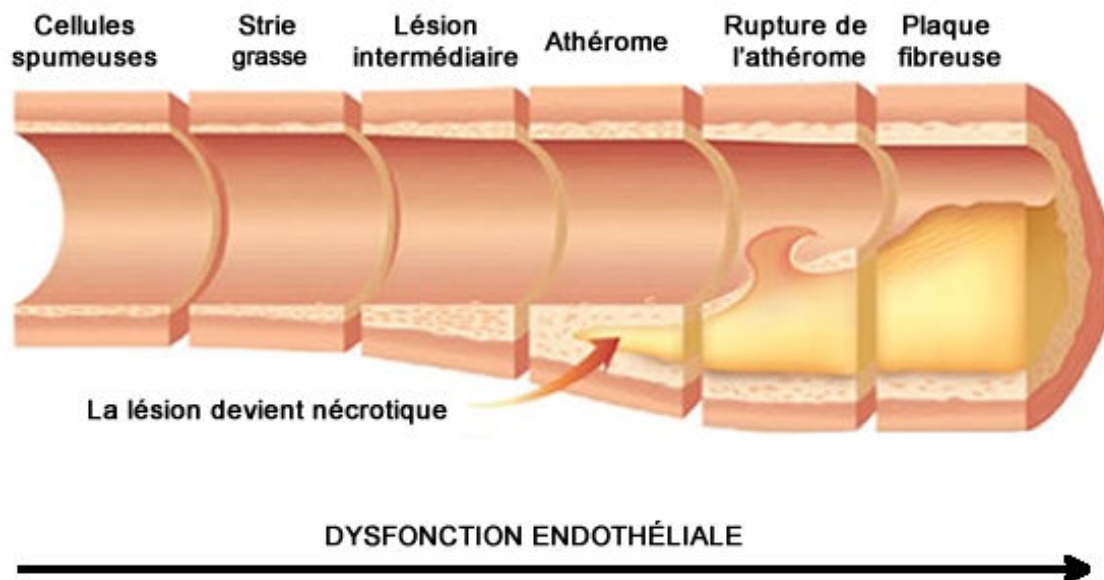


FIGURE 3.1 – Processus de développement d'athérosclérose par dysfonction endothéliale

L'inflammation de la paroi artérielle est considérée comme le dysfonctionnement du processus vasculaire artériel dont la cause est la *dysfonction endothéliale* (rôle central dans la régulation des phénomènes vasomoteurs et de la prolifération des cellules musculaires lisses de la paroi artérielle). Cette dysfonction est également considérée comme *l'athérosclérose*². L'athérosclérose est un facteur prédictif du risque cardiovasculaire chez les patients souffrant de maladies cardiovasculaires et chez les personnes apparemment en bonne santé. Les fonctions affectées par l'athérosclérose sont les suivantes :

- **Les fonctions cardiaques** : l'insuffisance cardiaque est une affection prolongée (chronique) qui peut apparaître soudainement. Cette affection apparaît lorsque le sang n'est pas pompé ou éjecté dans le cœur, mais également lorsque les muscles cardiaques sont rigides et ne se remplissent pas de sang facilement. Ces problèmes signifient que le cœur ne peut pas pomper de sang et oxygéner suffisamment le corps.
- **Les fonctions cérébrales** : un accident vasculaire cérébral est caractérisé en grande partie par une interruption soudaine du flux sanguin vers le cerveau ou par une hémorragie due à une rupture des veines cérébrales.
- **Les fonctions périphériques** : une affection implique une obstruction des artères périphériques, le plus souvent celles des jambes. La maladie artérielle périphérique survient très couramment chez les patients plus âgés et elle va souvent de pair avec

1. OMS, <http://www.who.int/fr/>

2. athérosclérose, <https://www.nhlbi.nih.gov/health/health-topics/topics/atherosclerosis/>

une maladie coronarienne. La plupart des lésions dues à la maladie périphérique apparaissent avant même les premiers symptômes.

- **Les fonctions rénales** : les maladies affectent les artères et les veines des reins qui provoquent une insuffisance rénale ou une hypertension artérielle.

3.2/ MÉTHODE DE CONCEPTION DE NOS ONTOLOGIES

Entre 2000 et 2010 les ontologies ont occupé une place importante dans l'ingénierie des connaissances [Yun11]. Plusieurs groupes de recherche ont proposé des méthodes de développement des ontologies, mais les domaines à modéliser sont si nombreux et si divers qu'il est impossible d'obtenir une seule méthode adaptée à tous les cas. La section 2.1.2 a introduit les méthodologies les plus utilisées.

Nous avons entre autres détaillé : *Ontology development 101* [Noy01] proposée par l'Université de Standord et *Methontology*³ [Cor03]. Ces deux méthodologies s'appliquent dans le cadre du développement d'ontologies à partir de zéro, de la réutilisation d'ontologies ou d'un processus de ré-ingénierie. Ces méthodologies comportent un ensemble établi de principes, de processus, de pratiques, de méthodes et d'activités utilisés pour le design, le développement, l'évaluation et l'implémentation [Gas09]. De telles méthodologies incluent des méthodes pour l'unification, le *reengineering*, la maintenance et l'évolution des ontologies. Nous nous appuyons sur la combinaison de ces deux méthodologies pour le développement de nos ontologies dédiées à l'organisation de connaissances dans le domaine des maladies vasculaires, du diagnostic lié à ces maladies et à la traçabilité des traitements et analyses du patient.

Ces méthodologies proposent un processus de développement qui s'appuie sur la construction d'un modèle conceptuel robuste et sur la détermination claire et consistante des exigences de l'ontologie à construire par :

- La détermination des besoins de l'ontologie (portée de l'ontologie, et ensemble des questions de compétences) ;
- La réutilisation des ontologies ou des méta-données existantes ;
- Le développement d'un modèle conceptuel ;
- La mise en œuvre du modèle conceptuel ;
- L'évaluation de l'ontologie.

Il est important de déterminer l'ensemble des questions de compétences [Noy01] car leur définition va nous aider pour la prise de décisions pendant le processus de développement de notre ontologie. *Ontology development 101* donne la définition suivante : « une ontologie est un modèle de la réalité et il n'y a pas qu'une seule façon correcte de modéliser son domaine, il y a toujours des alternatives viables. La meilleure solution dépend presque toujours de l'application que nous voulons mettre en place et des évolutions que nous anticipons ».

Les processus d'évaluation et de documentation sont des axes transversaux de ces méthodologies, ce qui implique leur présence continue dans chacune des étapes ci-dessus, mais pas toujours avec le même degré de formalité. Nous allons utiliser divers outils fournis par *Methontology* (schémas, diagrammes et descripteurs) pour documenter correctement le processus d'élaboration du modèle conceptuel. La Figure 3.2 illustre le

3. Methontology, <http://semanticweb.org/wiki/METHONTOLOGY>

processus de développement et de mise en œuvre de l'ontologie avec les outils utilisés à chaque étape.

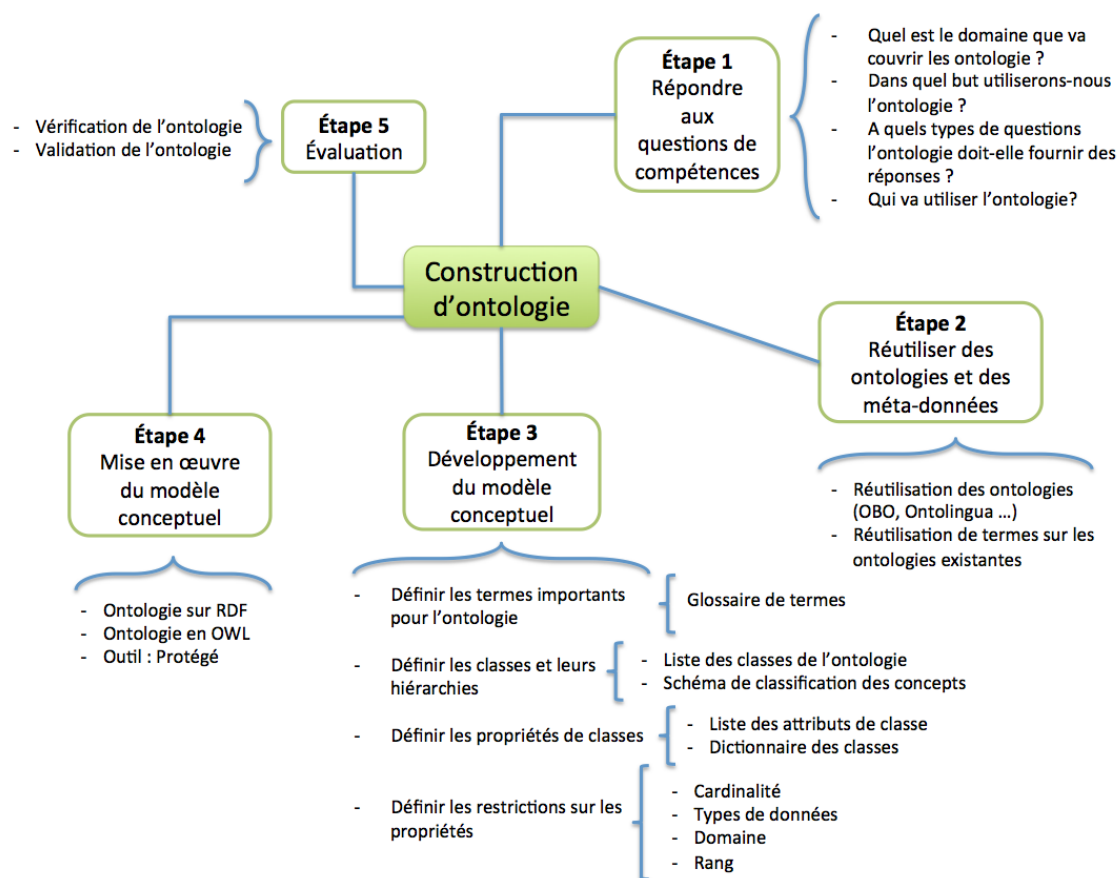


FIGURE 3.2 – Description de la méthode de développement de l'ontologie

3.2.1/ RÉPONDRE AUX QUESTIONS DE COMPÉTENCES

Nous commençons le développement de nos ontologies par la définition du domaine, l'apport et le degré de granularité de celles-ci. Pour cette tâche, il est suggéré d'utiliser des *questions de compétences*, c'est-à-dire, l'utilisation d'une liste de questions auxquelles le système ontologique doit répondre. Les questions de compétences et les réponses associées pour notre projet sont les suivantes :

Quel est le domaine que vont couvrir les ontologies ? Nos ontologies permettent de modéliser la structure du diagnostic des maladies vasculaires à partir de laquelle la communauté médicale peut collaborer pour un diagnostic et/ou une prise de décision de diagnostic. Les ontologies construites sont également utilisées pour permettre une surveillance médicale (la traçabilité des interventions ou études, l'historique clinique, les images d'explorations, les traitements et les résultats de diagnostic) d'un patient.

Dans quel but utiliserons-nous les ontologies ? Les ontologies facilitent la recherche d'informations et de concepts concernant les diagnostics dans le standard d'ima-

gerie médicale (DICOM)(cf. section 3.3.2) pour fournir aux médecins l'ensemble des informations relatives à un patient et son évolution clinique, tout en conservant la traçabilité du patient dans un objectif de travail collaboratif ultérieur. Pour l'acquisition de cette information, il est nécessaire d'inclure des concepts qui décrivent à la fois une maladie vasculaire, le diagnostic médical lié à la maladie et la collaboration médicale. Ces concepts donnent un guide médical sur l'évolution du patient. Chaque résultat doit exclure les éléments répétés dans les diagnostics qui peuvent créer de la confusion.

A quels types de questions les ontologies doivent-elles fournir des réponses ?

Dans le domaine du diagnostic médical sur les maladies vasculaires, il est nécessaire de pouvoir répondre aux questions suivantes : quelle maladie a le patient ? Quelles études ont été faites sur le patient ? Quel médecin ou quels médecins ont traité le patient ? Quelle est l'évolution du patient ? Combien et quelles sont les études physiologiques du patient ? Quels traitements a suivi le patient ? Quelle est l'historique clinique du patient ? Quelles institutions de santé ont traité le patient ? ...

Qui va utiliser les ontologies ? Les utilisateurs potentiels de ces ontologies sont les médecins qui participent de manière individuelle ou collaborative au diagnostic.

L'utilisation des questions de compétences est le départ de notre travail. Ces questions contraignent l'ingénieur des connaissances à un engagement sémantique, c'est-à-dire, d'explicitier clairement le sens de chacune des tâches de l'ontologie.

3.2.2/ PROCESSUS DE DÉVELOPPEMENT

Methontology [Cor03] propose de commencer par une tâche de planification pour identifier les tâches à réaliser, les corrections, le temps et les ressources nécessaires. Suivent les activités d'administration (contrôle et assurance de la qualité) et de support (acquisition de connaissances, intégration, évaluation, documentation et gestion de configuration) qui sont effectuées en parallèle avec les activités de développement (spécification, conceptualisation, formalisation, implémentation et maintenance) tout au long du cycle de vie de l'ontologie.

Après avoir spécifié le premier prototype, le modèle conceptuel est construit principalement en soutien de l'activité d'acquisition des connaissances. Au terme de la conceptualisation, les activités de formalisation peuvent être effectuées. Si aucun détail n'est détecté dans certaines de ces activités, il est possible de retourner à des activités antérieures pour apporter des modifications ou des améliorations.

L'activité de conceptualisation de l'ontologie mérite une attention spéciale dans l'organisation des tâches à réaliser. En principe, il est nécessaire d'avoir une perception informelle du domaine en utilisant un ensemble de représentations intermédiaires (tables, graphiques, schémas) qui peuvent être comprises par les experts du domaine et les développeurs d'ontologies. La Figure 3.3 montre les composants de l'ontologie (les concepts, les attributs, les relations constantes, les axiomes formels, les règles et les instances) construits durant chaque tâche.

Les tâches de conceptualisation de l'ontologie sont décrites ci-dessous :

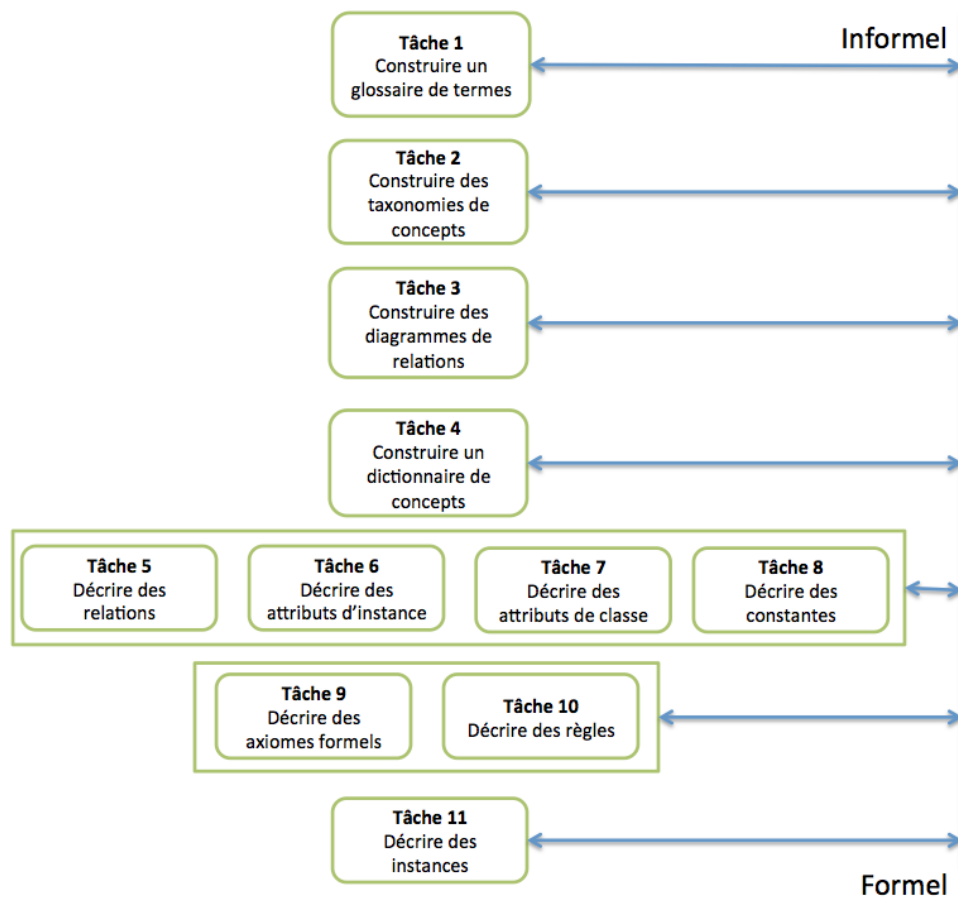


FIGURE 3.3 – Tâches de conceptualisation d'ontologie

Tâche 1 : construire le glossaire de termes. Le glossaire doit inclure tous les termes remarquables du domaine (les concepts, les instances, les attributs, les relations entre concepts, ...). Il doit également inclure les descriptions en langage naturel, les synonymes et les acronymes. Il est important de mentionner qu'au début, il peut exister divers termes qui se réfèrent au même concept. Ceux-ci doivent être identifiés et enregistrés comme des synonymes.

Tâche 2 : construire la taxonomie de concepts. Lorsque le glossaire de termes a un nombre important d'éléments, il constitue une taxonomie qui définit la hiérarchie entre les concepts. Pour réaliser une telle taxonomie, nous devons sélectionner les termes classifiés comme des concepts dans le glossaire de termes et construire une hiérarchie basée sur les quatre relations taxinomiques suivantes :

1. Sous-classe de : un concept C_1 est sous-classe d'un autre concept C_2 , si et seulement si toutes les instances de C_1 sont aussi des instances de C_2 .
2. Décomposition disjointe : une décomposition disjointe d'un concept C est une sous-partie de sous-classe de C qui n'a pas d'instance, c'est-à-dire, il peut y avoir des instances d'un concept C et ceux-ci ne sont pas des instances de l'un des autres concepts dans la décomposition.
3. Décomposition exhaustive : une décomposition exhaustive d'un concept C est un

ensemble de sous-classes de C qui couvre C et qui peut avoir des instances et des sous-classes communes, c'est-à-dire, il ne peut pas y avoir des instances du concept C qui ne sont pas instances d'au moins un des concepts de décomposition.

4. Partition : une partition d'un concept C est un ensemble de sous-classes de C qui ne partagent pas de cas commun, mais qui couvrent C , c'est-à-dire, il n'existe pas de cas de C qui ne sont pas des instances d'un des concepts de la partition.

À la fin, il faut évaluer la taxonomie créée pour vérifier qu'il n'existe pas d'erreurs, telles que l'héritage cyclique, l'erreur de partition, l'erreur sémantique ou la classification de concepts incomplète.

Tâche 3 : construire un diagramme de relations. Le but de ce diagramme est d'établir les relations entre les concepts d'une ou plusieurs taxonomies de concepts. Il faut vérifier que le diagramme ne contient pas d'erreur.

Tâche 4 : construire le dictionnaire de concepts. Le dictionnaire de concepts contient les concepts du domaine, leurs relations, leurs instances, leurs attributs de classes et leurs attributs d'instances. Les relations, attributs d'instances et attributs de classes sont liés à un concept, ce qui signifie que leurs noms peuvent être répétés dans différents concepts.

Tâche 5 : décrire les relations en détail. Il faut créer la table de relations dans laquelle est décrite de façon détaillée toutes les relations incluses dans le dictionnaire des concepts. Pour chaque relation, il faut spécifier le nom, les concepts sources et destinations, la cardinalité et la relation inverse.

Tâche 6 : décrire les attributs d'instances en détail. Il faut créer la table d'attributs d'instances et décrire en détail tous les attributs d'instances inclus dans le dictionnaire de concepts. Les attributs d'instances sont les attributs qui décrivent les instances d'un concept et leurs valeurs, celles-ci pouvant être différentes pour chaque instance du concept. Pour chaque attribut d'instance, il faut spécifier le nom, le concept auquel il appartient, le type de valeur, le rang de valeurs (dans le cas de valeurs numériques) et la cardinalité.

Tâche 7 : décrire les attributs de classe en détail. Il faut créer une table d'attributs de classe et décrire en détail tous les attributs de classe inclus dans le dictionnaire de concepts. Par chaque attribut de classe, il faut spécifier le nom, l'endroit où le concept est défini, le type de valeur, la valeur et la cardinalité.

Tâche 8 : décrire les constantes en détail. Il faut créer une table de constantes et décrire en détail chaque constante définie dans le glossaire de termes. Pour chaque constante, il faut spécifier le nom, le type de valeur, la valeur et l'unité de mesure (pour les constantes numériques).

Tâche 9 : décrire les axiomes formels. Il faut identifier les axiomes formels nécessaires à l'ontologie et les décrire avec précision dans une table. Pour chaque définition d'axiome formel, il faut spécifier le nom, la description, l'expression logique qui le décrit formellement (de préférence en utilisant la logique du premier ordre), les concepts, les attributs et les relations auxquelles l'axiome fait référence ainsi que les variables utilisées.

Tâche 10 : définir les règles. Il faut identifier quelles règles sont nécessaires dans l'ontologie et les décrire dans une table de règles. Pour chaque règle, il faut spécifier : le nom, la description, l'expression qui la décrit formellement, les concepts, les attributs, les relations auxquelles elles font référence et les variables utilisées dans l'expression. Pour la spécification des règles, il est suggéré d'utiliser la syntaxe suivante :

Si *< condition >* alors *< conséquences et actions >*

Tâche 11 : décrire les instances. Une fois que le modèle conceptuel de l'ontologie a été créé, il faut définir les instances remarquables qui apparaissent dans le dictionnaire de concepts dans une table d'instances. Pour chaque instance, il faut spécifier le nom, le concept auquel il appartient et les valeurs des attributs.

Ces tâches représentent la *normalisation*, la *formalisation* et l' *opérationnalisation* d'une ontologie (Figure 3.4)

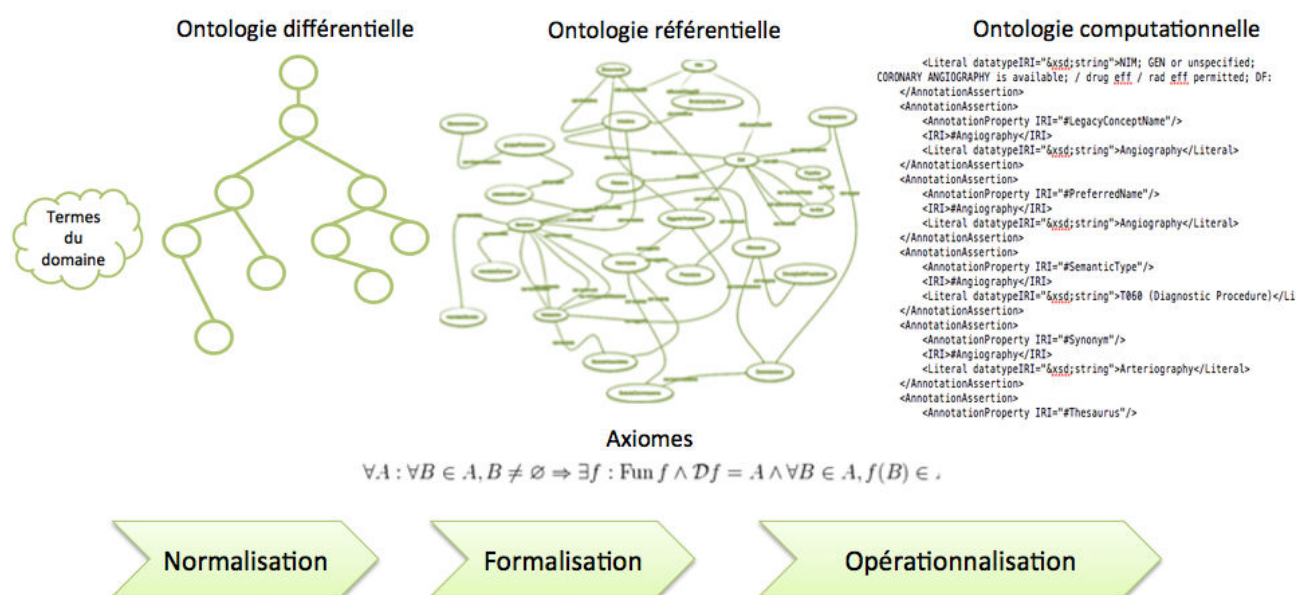


FIGURE 3.4 – Étapes générales de développement d'une ontologie

L'étape de normalisation sémantique a pour objectif de rendre explicite le sens des expressions linguistiques. Il s'agit, par ce processus, de créer des primitives du domaine, c'est-à-dire, d'identifier les notions élémentaires à partir desquelles l'ensemble des connaissances du domaine est construit. La langue est le média naturel de l'accès et de la diffusion des connaissances, comme les outils d'extraction terminologique (UMLS, ICD, MeSH, ...). Ces outils attribuent un sens aux termes grâce à la définition de caractéristiques sémantiques génériques et spécifiques. Ces caractéristiques permettent

non seulement de fixer le cadre interprétatif, en fonction de l'objectif que s'est donné l'ingénieur ontologique, mais également d'obtenir une primitive exploitable. Cela revient à associer aux termes une signification puisqu'ils sont décrits selon un certain point de vue, en l'occurrence celui de la tâche à réaliser. En pratique, l'ingénieur ontologique doit exprimer en langage naturel les identités (caractéristiques sémantiques génériques) et les différences (caractéristiques sémantiques spécifiques en opposition les unes avec les autres) que chaque terme entretient avec ceux qui lui sont proches. La structuration de ces termes, en fonction des identités et des différences qu'elles partagent avec leurs termes pères et leurs termes frères dans un arbre, permet de passer à *l'ontologie différentielle*.

L'ingénieur ontologique obtient à la fin de l'étape de normalisation une taxonomie de termes. La signification de chacune s'obtient de manière composée en parcourant les identités et les différences qui définissent l'ensemble des termes de l'arbre : du plus générique (le terme racine) au terme cible considéré. Autrement dit, la position d'un nœud dans l'arbre ontologique conditionne sa signification. Le processus de normalisation sémantique permet de passer d'un terme candidat à une notion dont le sens est invariable, et par conséquent, à une primitive représentant une connaissance du domaine à modéliser.

La normalisation des connaissances permet de formaliser les connaissances du domaine à représenter. Il s'agit de définir des concepts, et non plus des termes, selon une sémantique formelle et extensionnelle. Grâce à cela, les concepts pourront servir comme primitives dans un langage formel de représentation des connaissances. Il faut passer de la dimension linguistique et interprétative de la taxinomie à *l'ontologie référentielle* ou *l'ontologie formelle* composée de concepts et des significations. Selon la sémantique extensionnelle, les concepts sont liés à un ensemble de référents d'un ensemble d'objets du domaine. Cet ensemble est appelé l'extension du concept. L'ingénieur ontologique peut mettre en œuvre des opérations ensemblistes, telles que l'union, l'intersection, ... (cf. section 3.3.1.2), qui lui permettront de composer de nouveaux sens et donc de nouveaux concepts formels.

La structure de la hiérarchie ontologique n'est plus alors un arbre mais un graphe conceptuel puisque les extensions des concepts peuvent avoir un sous-ensemble commun. L'héritage multiple est alors possible (Figure 3.4). Cette étape de la méthodologie permet également de formaliser les relations qui existent entre les concepts en définissant leur parité et les ensembles d'extensions de concepts qu'elles relient.

L'étape d'opérationnalisation s'attache à informatiser l'ontologie référentielle dans un langage opérationnel de représentation des connaissances, en adoptant le formalisme des graphes conceptuels. En effet, un système informatique ne peut pas manipuler des concepts en fonction de leur interprétation sémantique. Il ne peut exploiter les concepts qu'en suivant les règles et les opérations associées. Ainsi, la sémantique qui permet à une machine d'utiliser des concepts réside dans la spécification informatique des opérations mathématiques que l'on peut faire sur ces concepts. Ces opérations peuvent être de plusieurs sortes en fonction du formalisme de représentation des connaissances choisis (RDF et OWL). Il s'agit ici de définir une *sémantique computationnelle* pour chaque concept de l'ontologie, c'est-à-dire que chaque concept est vu par la machine comme le résultat d'un ensemble d'inférences et de calculs. Cette étape marque le passage à *l'ontologie computationnelle*.

3.3/ DÉFINITION DE NOS DEUX PREMIÈRES ONTOLOGIES DANS LE DOMAINE VASCULAIRE

Le langage médical est caractérisé par un vocabulaire extrêmement riche et difficile à manipuler. Il n'y a pas de consensus établi sur la définition des termes employés. Les synonymes sont nombreux (plusieurs termes désignant le même concept) tandis que le même terme peut avoir plusieurs significations selon l'auteur ou le contexte (polysémie). Les textes médicaux sont donc souvent imprécis, ambigus d'autant qu'ils font un large usage d'abréviations et d'acronymes. Pour permettre une description et une communication efficaces et dépourvues d'ambiguïté, a fortiori pour un traitement automatique, un minimum de standardisation du langage est nécessaire.

Comme mentionné dans l'état de l'art, l'idée d'automatiser le raisonnement avec la connaissance formelle n'est pas nouvelle. Cependant, la vision de l'utilisation de la connaissance codée (ontologies) pour partager et échanger des connaissances entre les différentes personnes est relativement nouvelle et a suscité un intérêt considérable dans le développement des ontologies. Enfin, l'émergence du Web a conduit à la nécessité de partager systématiquement d'énormes quantités de données, ce qui constitue un facteur important pour l'intégration des ontologies dans les sciences de la santé par ordinateur. Même si l'infrastructure de *World Wide Web* permet aux humains de partager leurs connaissances (dans un format compréhensible par l'homme), sans ontologie, il n'existe aucun mécanisme efficace pour partager les connaissances entre les agents logiciels et des applications.

Pour commencer, il faut évidemment se demander quel est l'intérêt de nos ontologies. Dans notre cas, il est de pouvoir faire, un certain nombre de raisonnements basés sur la structure des ontologies et des relations entre les concepts ou notions. Ainsi, en dehors des relations de subsomption, nous avons modélisé le lien entre des symptômes ou des maladies et des spécificités des maladies vasculaires. Ces liens permettent de remonter à une interface des connaissances pour faciliter la prise collaborative de décision clinique : donc la collaboration entre les médecins distants.

L'ontologie de système vasculaire (OSV) et l'ontologie du diagnostic (OD) ont pour objectif de fournir un environnement pour l'intégration des ressources de diagnostic et l'interprétation des maladies présentes dans les diagnostics. Cette tâche est possible par la manipulation des images et des modules d'analyse et le traitement de ces images au format DICOM. Cet environnement ontologique se confronte à des applications cibles orientées vers le suivi de pathologies vasculaires (par exemples, accident vasculaire du cerveau ou du cœur, sclérose en plaques, ischémie, thrombose, infarctus, ...).

L'existence d'une telle pathologie diagnostiquée doit faciliter la réalisation de travaux de recherche multicentres dans le domaine vasculaire clinique. Ces travaux de recherches permettent la création de cohortes (étude médicale d'observation et analytique des maladies) de sujets diagnostiqués et facilitent l'accès à des sources cliniques via le web pour l'aide à la décision clinique. Les deux ontologies proposées vont composer une nouvelle chaîne optimale de traitement et de diagnostic collaboratif (cf. section 3.4).

3.3.1/ ONTOLOGIE DU SYSTÈME VASCULAIRE

L'objectif du développement d'une ontologie dans le domaine vasculaire est l'intégration d'informations médicales à large échelle, avec un accent particulier sur *l'aide à la collaboration diagnostique pour la prise de décisions cliniques*. Le domaine d'application est celui des maladies vasculaires ce qui comprend les maladies du cerveau, du cœur, rénales et périphériques. Cette ontologie a pour but l'indexation automatique et manuelle des connaissances, en fournissant un soutien direct ou indirect aux utilisateurs que sont les médecins spécialistes qui travaillent de manière collaborative.

Le développement d'une ontologie *à partir de rien* dans le domaine biomédical nécessite l'aide d'experts et l'analyse de textes. Cette tâche exige un grand effort de conceptualisation et beaucoup de temps pour l'éditer (avec l'éditeur d'ontologie comme Protégé). La méthode Bottom-up [Bla02, Kri12] vise à identifier les concepts et les relations à partir de documents du domaine et de questions posées à l'utilisateur dans les limites du domaine intéressé et à les transformer en une ontologie. Le *corpus* est l'expression des connaissances à travers une langue de spécialité (vocabulaire, terminologie et sémantique spécifique) qui est utilisé pour échanger dans le domaine. L'obtention du corpus s'est fait à travers un relevé des informations et connaissances (état de l'art, DICOM, documents, experts, ...). L'objectif est l'obtention d'un corpus comprenant des connaissances dont on aura besoin pour nourrir la représentation ontologique. L'étape précédente d'identification des concepts et des relations propose les candidats puis il faut réaliser une normalisation sémantique pour éviter les problèmes de langage naturel.

3.3.1.1/ COMPOSANTS DE L'ONTOLOGIE

La modélisation des connaissances d'un domaine avec les logiques de description (LD) peut se réaliser en deux niveaux séparant les représentations de connaissances intentionnelles et extensionnelles. Le premier est un niveau terminologique ou *TBox*, qui décrit les connaissances générales d'un domaine alors que le second est un niveau factuel ou *ABox*, et représente une configuration précise [Zom10]. Ainsi nous pouvons définir plus précisément :

- La *TBox* (T pour terminologie) décrit les connaissances générales d'un domaine et contient les déclarations des primitives conceptuelles, organisées en concepts et relations, qui peuvent également être combinés grâce à des constructeurs, pour former de nouveaux concepts. Ces déclarations décrivent les propriétés des concepts et des relations, elles constituent donc une définition intentionnelle des connaissances.
- Une *TBox* (A pour assertion) décrit les connaissances factuelles d'un domaine et représente une configuration précise. Elle contient les déclarations d'individus, instances des concepts qui ont été définis dans la Tbox. Plusieurs Abox peuvent être associées à une même Tbox ; chacune représente une configuration constituée d'individus et utilise les concepts et rôles de la TBox pour l'exprimer.

Cette approche nous a permis : de déterminer pour la TBox s'il existe une interprétation quelconque pour laquelle un concept a une instance et pour la ABox de déterminer si un individu particulier est ou n'est pas une instance d'un concept (instanciation).

Nous pouvons envisager, par exemple, une classe ontologique appelée *HypertensivePatient*. Cette classe se composera de l'ontologie initiale (TBox). Quand nous commençons le remplissage de la classe avec les instances sur les critères d'adhésion, nous créons

effectivement la base de connaissances (ABox).

Les principaux composants d'une ontologie sont les concepts (classes), les relations, les instances et les axiomes. Un concept représente un ensemble d'entités dans un domaine. Par exemple, un « vaisseau sanguin » est un concept dans le domaine de la cardiologie. Les concepts sont l'objectif principal de la plupart des ontologies. Les concepts peuvent être divisés en deux catégories :

- Concepts primitifs : ce sont ceux qui ont des conditions nécessaires (en termes de classes avec des propriétés existantes) pour la relation de membre de classe. Par exemple, si nous affirmons qu'il est nécessaire que le *Patient* soit une *Personne*. Évidemment, la relation inverse n'existe pas.
- Concepts définis : ce sont ceux dont la description est à la fois nécessaire et suffisante pour qu'un terme soit un membre de la classe. Par exemple, GALEN⁴ définit l'hypertension légère (MildHypertension) comme suit :

$$\text{MildHypertension} = \text{Hypertension AND hasSeverity MildSeverity}$$

Sur la base de l'exemple ci-dessus (MildHypertension), toute entité qui remplit la condition droite de l'équation implique d'être membre de la classe MildHypertension. Globalement, il existe deux types de relations : relation de taxonomie et relation associative.

Les taxonomies qui organisent des concepts (sous-super-concept) dans une structure hiérarchique. Les relations de hiérarchie les plus utilisées sont : les *relations de spécialisation* connues sous le nom « is-a » ou « is-a-kind-of » (par exemple MildHypertension is-a-kind-of Hypertension), et les *relations partitives* décrivent les concepts qui font parties d'autres concepts (par exemple Heart hasComponent HeartValve).

Les relations associatives lient les concepts dans l'arbre structurel. Les relations plus communes sont : les *relations nominatives* qui décrivent les noms des concepts (par exemple Disease hasName CongestiveHeartFailure), les *relations locatives* décrivent l'emplacement d'une notion ou d'un concept l'un par rapport à l'autre (par exemple HeartValve hasPhysicalLocation Heart), et les *relations associatives* qui représentent les propriétés qui peuvent être attribuées à un concept (par exemple CongestiveHeartFailure playsPathologicalRole Disease).

Les relations peuvent également être hiérarchiques. Par exemple, la relation *hasName* peut être subdivisée en *hasDiseaseName* et *hasPatientName*. Ces relations ont également des fonctionnalités (comme la transitivité et la cardinalité) qui capturent des connaissances supplémentaires sur les relations entre les concepts.

3.3.1.2/ FORMALISMES POUR LA REPRÉSENTATION DES CONNAISSANCES

Représenter des connaissances propres à un domaine consiste à décrire et à coder les éléments de ce domaine pour qu'une machine puisse les traiter, raisonner et résoudre des problèmes particuliers [Zom10]. Il faut donc savoir : (1) exprimer ces connaissances à l'aide d'un langage formel de description des connaissances et (2) les manipuler, c'est-à-dire être capable de faire un certain nombre d'opérations dessus (modifier, compléter, déduire de nouvelles connaissances, ...) à l'aide de mécanismes définis opérant sur les différents éléments de la représentation.

4. GALEN <http://www.ncbi.nlm.nih.gov/pubmed/7889770>

Il existe un certain nombre de formalismes dans le domaine de la représentation des connaissances : les graphes conceptuels [Cao10] et les logiques de description [Pit13]. L'une comme l'autre permettent de représenter des ontologies, des propriétés, et de mettre en œuvre des mécanismes d'inférence.

Le graphe conceptuel de cette ontologie du système vasculaire (Figure 3.5) se décompose en deux parties :

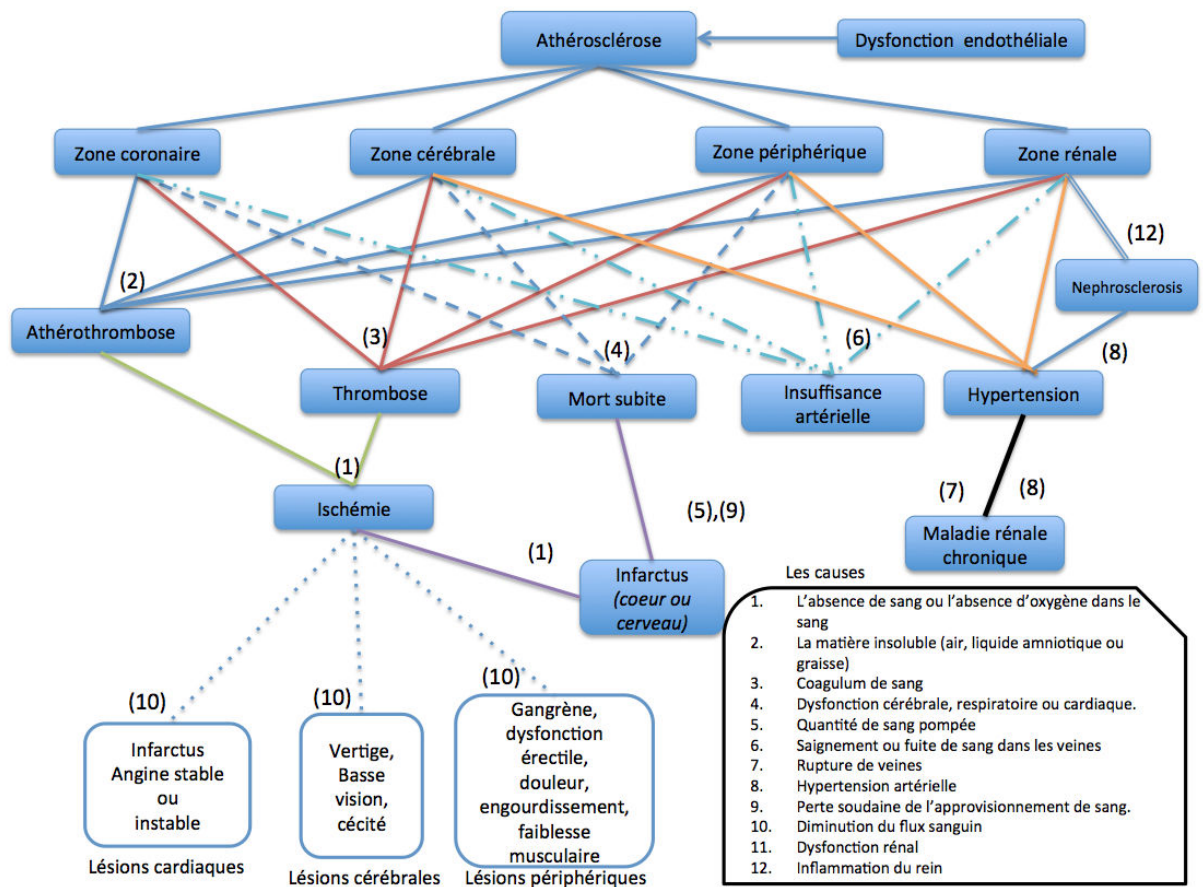


FIGURE 3.5 – Graphe conceptuel général du système vasculaire

- Une partie terminologique dédiée au vocabulaire conceptuel des connaissances à représenter (les types de concepts T_c , les types de relations T_r et les instances M des types de concepts). La partie relative à la terminologie correspond à la représentation du modèle conceptuel mais intègre également des connaissances sur la hiérarchisation des types de concepts et de relations. Ces trois ensembles constituent le support qui va régir l'ensemble des graphes conceptuels portant sur un même domaine de la connaissance. Un support se définit ainsi : $S = (T_c; T_r; M)$, les ensembles T_c , T_r et M étant disjoints et les ensembles T_c et T_r étant partiellement ordonnés.
- Une partie dédiée à la représentation des assertions du domaine des maladies vasculaires.

Les logiques de description aussi appelées Logique Descriptive (LD) sont des langages formels permettant de représenter des propriétés pour des ensembles de connaissances

terminologiques. LD est le nom d'une famille de formalismes pour la représentation de connaissances [Pit13].

Un attribut que distingue DL est le support au « raisonnement » (extraction de connaissances implicites de la connaissance explicite représentée dans une base de connaissances), comme la classification des concepts et des individus qui sont produits dans les applications du système de traitement d'information. Le Tableau 3.1 présente les règles de syntaxe de base pour la Logique Descriptive par exemple entre deux concepts C et D selon le rôle ou restriction R .

Symbole	Description	Exemple	Interprète
\top	Tous les noms de concepts	\top	Haut – Représente le concept plus général
\perp	Concept vide	\perp	Bas – Représente le concept moins général
\sqcap	Intersection ou conjonction des concepts	$C \sqcap D$	C et D
\sqcup	Union ou disjonction des concepts	$C \sqcup D$	C ou D
\neg	Négation ou complément des concepts	$\neg C$	Non C
\forall	Restriction universelle	$\forall R.C$	Tous R -successeurs sont dans C
\exists	Restriction existentielle	$\exists R.C$	Un R -successeur existe dans C
\sqsubseteq	Concept d'inclusion	$C \sqsubseteq D$	Toutes les instances de C sont dans D
\equiv	Concept d'équivalence	$C \equiv D$	C est équivalent à D

TABLE 3.1 – Règles de syntaxe de base pour LD

Par exemple si nous supposons la définition du concept « patient » comme une « personne qui a une maladie », selon le Tableau 3.1, la syntaxe de LD pour cette déclaration est la suivante :

$$Patient \equiv Person \sqcap \exists hasDisease.T$$

Les membres individuels d'une classe peuvent être spécifiés explicitement (par exemple Jean is-a Patient), les classes peuvent également être définies de façon plus générale en termes d'axiomes. Ces axiomes décrivent les conditions nécessaires ou conditions suffisantes concernant les individus qui sont membres d'une classe. Par exemple, un patient avec une pathologie cardiaque peut être défini comme suit :

$$HeartPatient(p) \equiv Patient \sqcap \exists hasDisease(p, d)$$

$$HeartDisease(d)$$

$$HeartDisease \sqsubseteq Disease$$

Où d affirme une maladie cardiaque *HeartDisease* qui est elle-même une sous-classe d'une maladie *Disease*. La règle $\exists hasDisease(p, d)$ est aussi appelée une restriction de propriété (dans cet exemple, une propriété « existentielle » dans le tableau 3.1). Il existe d'autres types de restrictions des propriétés, telles que les restrictions universelles, déclarant que toutes les valeurs d'une propriété doivent appartenir à certaines restrictions de classes et de cardinalité, et déclarant que l'instance doit avoir un certain nombre de valeurs distinctes pour une propriété.

L'utilisation d'axiomes pour la définition des classes et des propriétés permet de déterminer automatiquement les contradictions au sein d'une ontologie qui est appelé vérification de cohérence [Zom10]. Dans l'exemple précédent si nous avons p qui est déclaré comme une instance de *HeartPatient* mais qu'il n'est pas lié à une instance de *HeartDisease* pour la propriété existentielle $\exists hasDisease$, alors un système d'inférence (connu sous le nom OWL raisonneur) peut détecter cette incohérence automatiquement.

La nature combinatoire de OWL-DL permet à de nouveaux concepts d'être créés à partir de la combinaison de concepts existants et qui sont automatiquement placés dans la hiérarchie ontologique. Les services de raisonnement fournis par OWL reasoners (logiciels d'application qui peuvent faire des déductions logiques basées sur des axiomes OWL) ont de nombreuses applications qui sont spécifiquement adaptées pour les systèmes de classification dans le domaine biomédical.

La terminologie médicale utilisée dans le domaine de la Cardiologie et la Neurologie est caractérisée par un grand nombre de termes, qui sont décrits comme des artefacts linguistiques, lesquels relient les différents sens ou significations aux entités. Ces terminologies ont des finalités bien établies (par exemple la récupération d'information dans des documents, l'enregistrement de statistiques de mortalité et de morbidité, la facturation des services de santé, ...). La plupart de ces terminologies n'utilisent pas des descriptions formelles, elles définissent les termes à travers des expressions du langage humain et par des associations explicites entre les termes par des relations informelles, également proches du langage humain. Ils sont formés avec des racines grecques et latines, préfixes et suffixes ce qui vise à simplifier la langue, à trouver de la précision dans le sens des mots et à faciliter les échanges scientifiques entre les pays de différentes langues.

Les maladies vasculaires sont plus présentes au niveau du cœur. Un exemple simple sur les fonctions du cœur est :

- *Le cœur est une pompe à sang, comme concept, il est composé de deux pompes distinctes : le côté droit qui pompe le sang des poumons et le côté gauche qui pompe le sang des organes périphériques. Chacune de ces parties séparées du cœur est composée d'un atrium et d'un ventricule. Ces deux grosses artères sont connectées à un système fermé (vaisseaux, vaines, capillaires) de distribution vasculaire. Si le cœur ou le système vasculaire souffre d'une insuffisance de sang, alors plusieurs facteurs physiologiques d'un patient peuvent se développer comme des dysfonctions ou des maladies. Par exemple : la cardiopathie, la sténose, l'insuffisance cardiaque, l'embolisme, la thromboses, l'accident vasculaire cérébral, la mort subite, ...*

Selon l'information précédente, la base de connaissances est composée de deux éléments : (1) le traitement de la terminologie, qui dispose d'un ensemble d'axiomes et des déclarations qui décrivent la structure du domaine (logique de description TBox) ; (2) le traitement des déclarations des éléments ou des objets (logique de description ABox). Autrement dit, l'ABox contient des connaissances intentionnelles sur le domaine

d'intérêt et le TBox est un modèle de base de déclarations des concepts et cette base est constituée d'un ensemble d'assertions et d'inclusions entre les concepts. Les composantes de base TBox et ABox sont présentées dans la Table 3.1. La déclaration ci-dessous est l'expression du concept de la cardiopathie valvulaire (*Valvular Heart Disease*), qui se traduit comme un élément du concept cœur (*Heart*) et du concept maladie cardiaque (*Heart Disease*).

$$\text{Valvular Heart Disease} \equiv \text{Heart} \sqcap \text{Heart Disease}$$

Selon la déclaration suivante, pour qu'il y ait la référence à l'artère coronaire, l'élément doit appartenir au concept de *Heart* et non au concept de *Valvular Heart Disease*.

$$\text{Coronary Artery} \equiv \text{Heart} \sqcap \neg \text{Valvular Heart Disease}$$

Pour l'ABox il est possible de faire deux types de déclarations différentes : la déclaration de concepts et la déclaration de rôles, où la construction de la connaissance extensionnelle prend place quand les concepts et les rôles sont des déclarations affirmatives (assertives) sur les éléments. La déclaration de concepts : *Valvular Heart Disease (Stenosis)* signifie que l'élément *Stenosis* appartient au concept *Valvular Heart Disease*. La déclaration de rôles : *Heart Diseases (Stenosis, Right Coronary Artery)* signifie que l'élément *Stenosis* se rapporte à l'élément *Right Coronary Artery* à travers le rôle de *heart_diseases*. Il est possible de faire différentes logiques de description, chacune étant spécifique à une situation particulière. Les classements suivants montrent les services typiques d'inférence dans TBox et ABox :

Insatisfiabilité de l'élément. L'insatisfaction des concepts est vérifiée si une description du concept ne peut pas avoir d'instances en raison d'incohérence ou de contradiction dans le modèle. Nous présentons ci-dessous un exemple d'insatisfiabilité dans le concept de *Cardiovascular System*.

$$\text{Valvular Heart Disease} \equiv \text{Heart} \sqcap \text{Heart Disease}$$

$$\text{Coronary Artery} \equiv \text{Heart} \sqcap \neg \text{Valvular Heart Disease}$$

$$\text{Cardiovascular System} \equiv \text{Coronary Artery} \sqcap \text{Valvular Heart Disease}$$

Vérification de classification. La vérification de classification entre deux descriptions de concepts peut être compris comme : un concept *C* et un concept *D*, où *C* contient *D* et lorsque l'ensemble des objets qui sont des instances de *D* sont également un sous-ensemble d'objets qui sont eux-mêmes des instances de *C*. Un exemple pratique de ces concepts est le concept *Heart Valve Diseases*, lequel est composé des instances du concept *Valvular Heart Disease*. Il y a de plus une relation entre la propriété *heart_diseases* et le concept de *Heart*.

$$\text{Valvular Heart Disease} \equiv \text{Heart} \sqcap \text{Heart Disease}$$

$$\text{Heart Valve Diseases} \equiv \text{Valvular Heart Disease} \sqcap \exists \text{heart_diseases.Heart}$$

L'équivalence entre les instances. Affirmer que les concepts C et D sont équivalents ($C \equiv D$), revient à dire qu'ils ont la même instance. Un exemple est le concept de *Right Coronary Artery* et *Electrical Conduction Disorders of the Heart*.

$$\text{Valvular Heart Disease} \equiv \text{Heart} \sqcap \text{Heart Disease}$$

$$\text{Right Coronary Artery} \equiv \text{Heart} \sqcap \neg \text{Valvular Heart Disease}$$

$$\text{Electrical Conduction Disorders of the Heart} \equiv \text{Heart} \sqcap \neg \text{Heart Disease}$$

Vérification de consistance et d'inférence en ABox. Le modèle ABox est consistant s'il existe une instance qui est vraie à la fois pour ABox et pour TBox. Par exemple, le concept *Valvular Heart Disease* contient *Stenosis* (la maladie sténose est présente dans la valve du cœur). Comme le concept *Valvular Heart Disease* est formé par les concepts de *Heart* et *Heart Diseases*, *Stenosis* fera également partie de ces concepts : *Valvular Heart Disease* étant un concept consistant dans le modèle ABox.

Le modèle TBox :

$$\text{Valvular Heart Disease} \equiv \text{Heart} \sqcap \text{Heart Disease}$$

Le modèle ABox :

$$\text{Valvular Heart Disease} (\text{Stenosis})$$

Vérification de consistance :

$$\text{Valvular Heart Disease} (\text{Stenosis}) \equiv \text{Heart} (\text{Stenosis}) \sqcap \text{Heart Disease} (\text{Stenosis})$$

Vérification de la procédure pour les éléments. Il s'agit de la vérification d'un élément donné comme étant une instance d'un concept spécifique. Un exemple d'une telle inférence est que l'élément *Stenosis* fait partie du concept *Valve Diseases of the Heart* : ce qui est explicite, c'est que la *Stenosis* a une relation directe avec le concept de *Valvular Heart Disease* par la propriété *heart_diseases* avec *Right Coronary Artery*. Nous avons comme résultat :

Le modèle TBox :

$$\text{Valvular Heart Disease} \equiv \text{Heart} \sqcap \text{Heart Disease}$$

$$\text{Heart Valve Diseases} \equiv \text{Valvular Heart Disease} \sqcap \exists \text{heart_diseases}.\text{Heart}$$

Le modèle ABox :

$$\text{Valvular Heart Disease} (\text{Stenosis})$$

$$\text{Coronary Artery} (\text{Right Coronary Artery})$$

$$\text{heart_diseases} (\text{Stenosis}, \text{Right Coronary Artery})$$

De cette information, nous pouvons dire que l'élément *Stenosis* appartient au concept de *Heart Valve Diseases*.

$$\text{Heart Valve Diseases}(\text{Stenosis}) \equiv \text{Valvular Heart}$$

$$\text{Disease}(\text{Stenosis}) \sqcap \exists \text{heart_diseases.Heart}$$

Retour d'élément. Cette tâche consiste à trouver le concept le plus spécifique sur lequel l'élément est une instance. Dans le cas de l'élément *Stenosis*, la détermination du concept le plus spécifique (concept le plus bas de la hiérarchisation) est rendue possible grâce au modèle suivant :

Le modèle ABox :

$$\text{Valvular Heart Disease}(\text{Stenosis})$$

$$\text{Heart Valve Diseases}(\text{Stenosis})$$

Retourne l'élément :

$$\text{Stenosis} \rightarrow \text{Valvular Heart Disease}$$

Exécution d'identification sur la base de connaissances. Ce service identifie dans la base de connaissances les éléments qui sont des instances d'un concept donné.

Le modèle ABox :

$$\text{Valvular Heart Disease}(\text{valvular insufficiency}) \sqcup \text{Valvular Heart Disease}(\text{Stenosis})$$

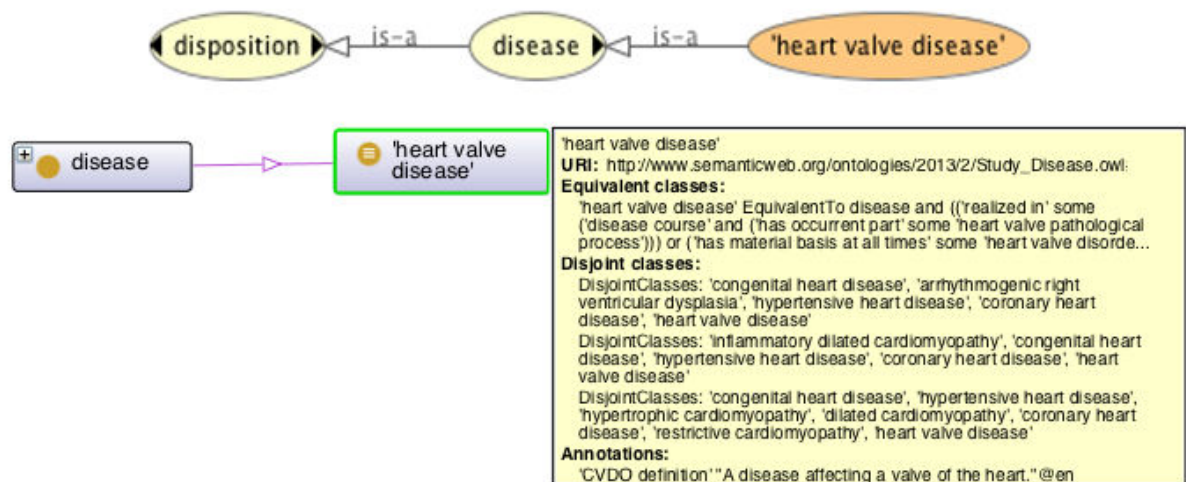
$$\text{Coronary Artery}(\text{Left Coronary Artery})$$

Résultat :

$$\text{Valvular Heart Disease} \rightarrow \text{Valvular Insufficiency}, \text{Stenosis}$$

$$\text{Coronary Artery} \rightarrow \text{Left Coronary Artery}$$

Il est important de prendre en considération que le développement d'une ontologie dans le domaine médical est aussi basé sur le raisonnement ontologique, lequel est utile et nécessaire. Ce type de logique de description ontologique soutient la prise de décision. La logique de description nous permet de décrire l'ontologie en langage de description (OWL) qui est utilisé pour définir la manière d'organiser les concepts et les rôles (TBox) et le système de spécification des propriétés sur les éléments (ABox). La Figure 3.6 montre la description graphique du concept *Heart Valve Disease*.

FIGURE 3.6 – Classe OWL du concept « *heart valve disease* » (pathologies du cœur)

3.3.1.3/ DÉVELOPPEMENT DE L'ONTOLOGIE

Le développement d'une ontologie médicale spécialisée *à partir de rien*, avec des connaissances d'experts ou de l'analyse de textes nécessite un énorme effort de conceptualisation et un long temps d'édition avec des éditeurs d'ontologies tels que *Protégé-2000*, qui possèdent déjà des thésaurus et terminologies standards.

Par conséquent, notre choix a été de s'appuyer sur les ressources existantes telles que le meta-thésaurus UMLS (*Unified Medical Language System*) et thésaurus MeSH (*Medical Subject Headings*), les deux utilisant l'indexation de documents. Par exemple, le système *Medline* indexe des articles scientifiques biomédicaux avec les concepts du MeSH.

Le thésaurus MeSH est un regroupement sémantique particulier. Les concepts utilisés ne sont pas de haut niveau de conceptualisation, mais il y a des descripteurs et des entités étroitement liés. Depuis que le thésaurus MeSH fournit des informations riches sur les concepts et leurs relations, des efforts ont été faits pour migrer ce thésaurus en OWL.

Pour représenter le lien entre les documents et les concepts de MeSH, les documents sont des instances de classes qui peuvent être des classes simples ou des catégories agrégées par les opérateurs d'union ou d'intersection, par lesquels le document sera indexé. Plusieurs ontologies médicales sont conçues pour la récupération de l'information et des annotations sémantiques. Pour cette raison les concepts dans MeSH doivent contenir tous les types possibles de vocabulaires utilisés dans le domaine. Notre ontologie a été créée en grande partie avec les concepts de MeSH en choisissant des concepts de 11 hiérarchies sur le domaine des maladies vasculaires. Ces hiérarchies sont identifiées pour un nom commun et le numéro de l'arbre :

- Pathologic Processes (C23.550),
- Vascular Diseases (C14.907),
- Vascular Malformation (C14.240.850, C16.131.240.850),
- Cerebrovascular Disorders (C10.228.140.300, C14.907.253),
- Intracranial Arteriosclerosis (C10.228.140.300.510.800, C14.907.137.126.372, C14.907.253.560.350),
- Cardiovascular Diseases (C14),

- Cardiovascular Diagnostic and Techniques (E01.370.370),
- kidney disease (C12.777.419, C13.351.968.419),
- Peripheral Artery Diseases (C14.907.137.126.307.500, C14.907.617.671),
- Atherosclerosis (C14.907.137.126.307),
- Embolism and Thrombosis (C14.907.355).

Certains éléments d'information extraits de MeSH comme des définitions, des commentaires, des termes, des relations (« is-a » et « see-also »), des références de concepts (id) ont également été intégrés dans l'ontologie du système vasculaire. L'enrichissement du vocabulaire a été conçu en deux étapes : (1) d'abord l'extraction des termes de MeSH et UMLS qui correspondent aux concepts de base en anglais, lesquels sont liés aux dysfonctionnements vasculaires ; (2) ensuite l'examen de textes en anglais du domaine des maladies vasculaires liées à l'athérosclérose.

```
<owl:Class rdf:ID="CVD0_0000039">
  <rdfs:label xml:lang="en">infarction</rdfs:label>
  <owl:equivalentClass>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <rdf:Description rdf:about="« #OGMS_0000061"/>
        <owl:Restriction>
          <owl:onProperty rdf:resource="#BF0_0000066"/>
          <owl:someValuesFrom rdf:resource="#OGMS_0000061"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </owl:equivalentClass>
  <CVD0_0000419 xml:lang="en">A necrosis of a tissue caused by a local lack
of oxygen, due to an obstruction of the tissue's blood supply.</CVD0_0000419>
</owl:Class>
```

FIGURE 3.7 – Classe OWL du concept Infarctus comme pathologies du cœur et du cerveau

Rappelons que nous travaillons avec les cardiologues mexicains de Morelia (en particulier pour évaluer et valider l'ontologie de la pathologie), et nos ontologies ont donc été conçues en anglais afin d'être également utilisables en Europe.

L'ontologie extraite du thésaurus MeSH pour le domaine vasculaire contient 525 classes OWL, chacune correspondant à un descripteur MeSH. La première étape de l'enrichissement du vocabulaire en langue anglaise a été d'augmenter de 7 320 le nombre de nouveaux termes, 2 500 étant des synonymes et les autres sont des variantes lexicales (Figure 3.9). Le concept est représenté par son ID provenant du MeSH et un terme appelé aussi terme préféré (preferred term). Nous avons associé dans cette phase Jean-Denys Maréchal, élève ingénieur en Génie Biomédical (École ISIFC) que j'ai encadré en stage de fin d'étude R & D.

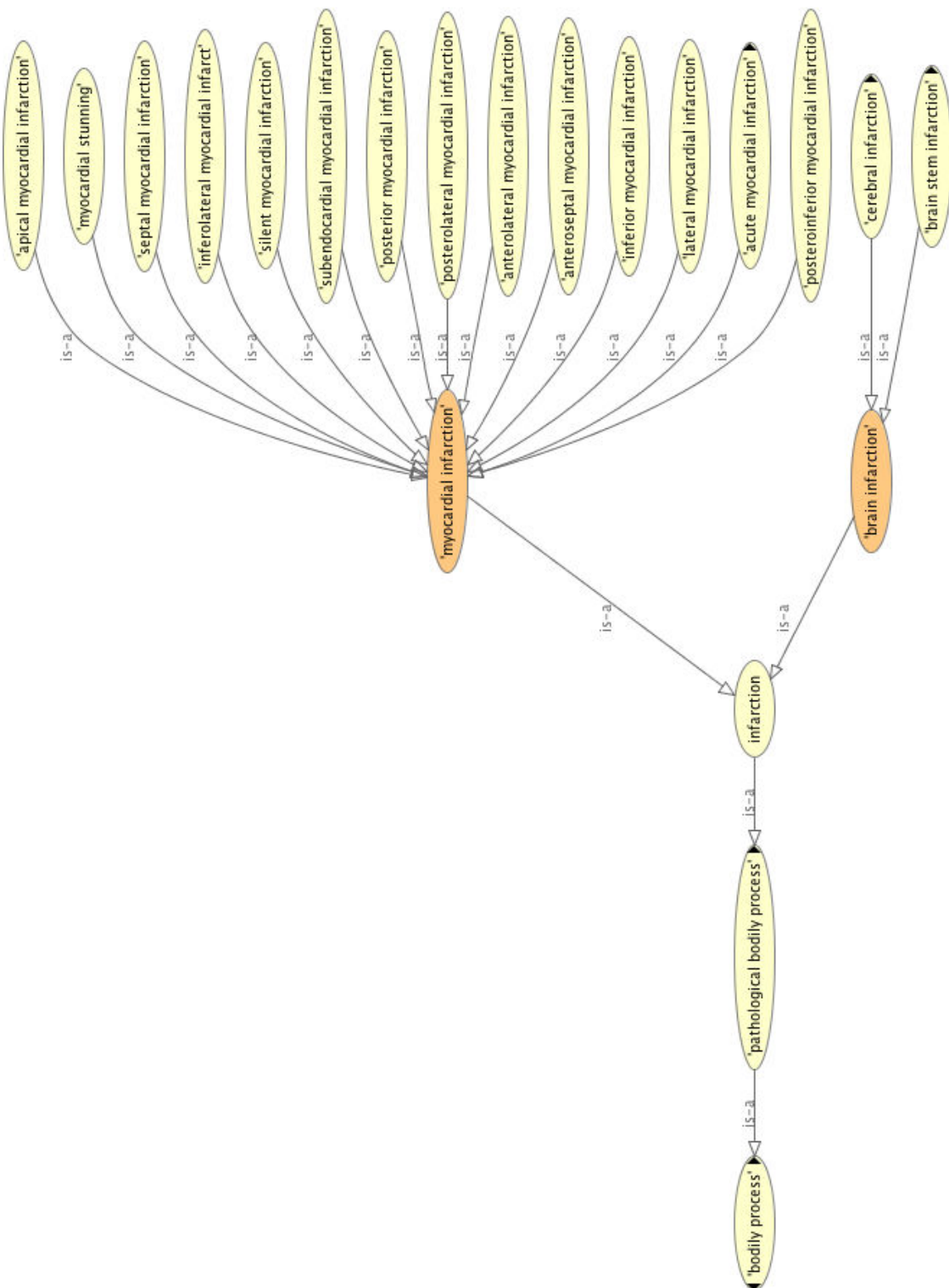


FIGURE 3.8 – Modèle spécifique du concept Infarctus

Un exemple d'une classe OWL est donné ci-dessous pour le concept Infarctus (Infarction) des pathologies du cœur (Figure 3.7 et Figure 3.8) qui est la sous-classe hiérarchique de *Pathological Bodily Process*. L'ontologie complète du système vasculaire (Figure 3.10) correspond à la modélisation des maladies du cœur, du cerveau, du système périphérique et rénal.

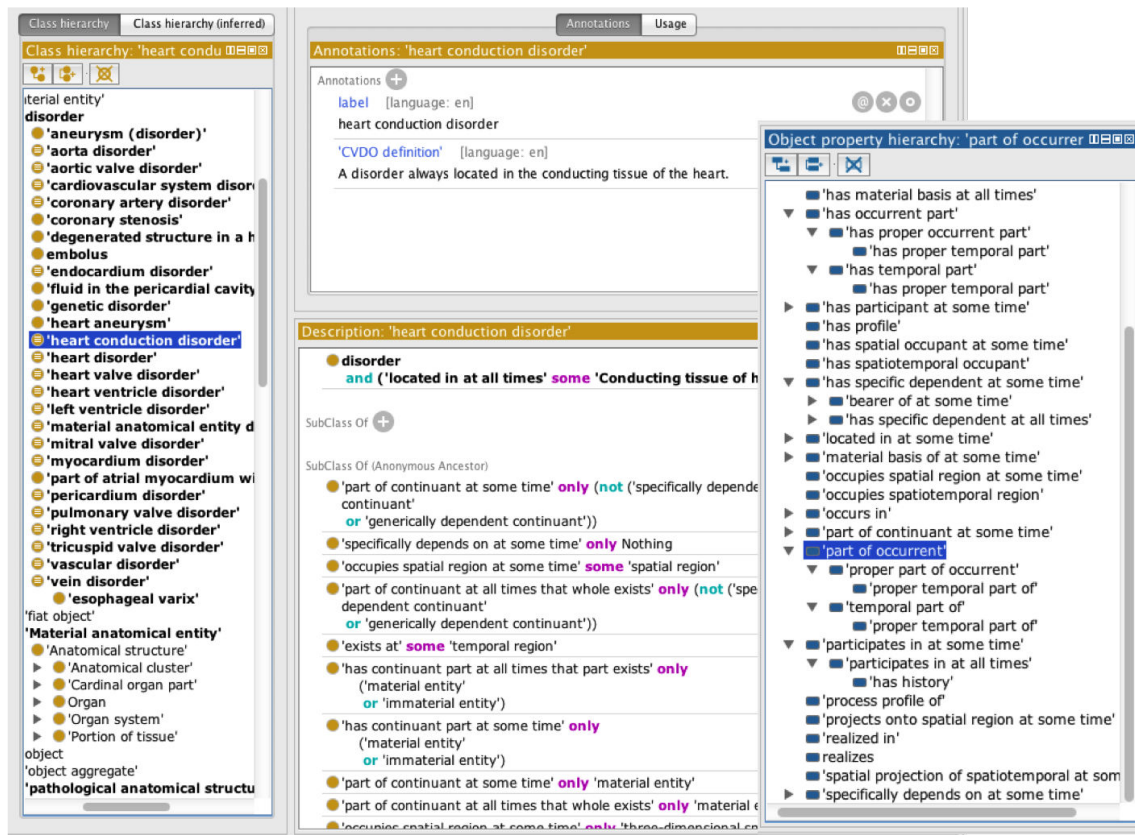


FIGURE 3.9 – Implémentation dans protégé de l'ontologie du système vasculaire

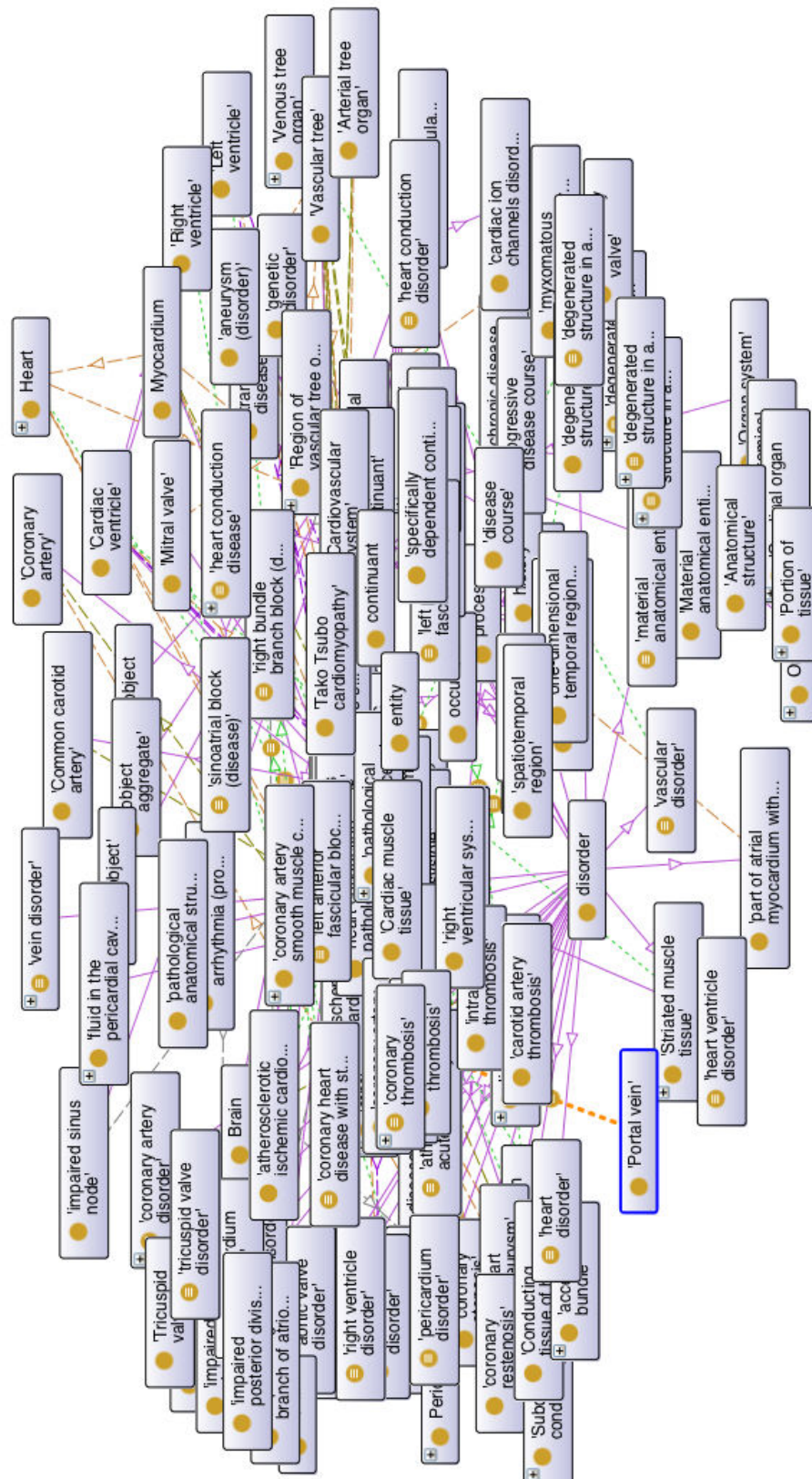


FIGURE 3.10 – Ontologie du système vasculaire

3.3.2/ ONTOLOGIE DU DIAGNOSTIC

Les ontologies ont également un intérêt dans les applications de diagnostic, surtout pour l'aide aux décisions cliniques. Les ontologies dans le domaine du diagnostic peuvent être utilisées pour représenter la connaissance globale des patients (par exemple données des patients, les diagnostics, les prescriptions, les résultats de tests, les imageries, ...) [Bla09]. Le diagnostic des maladies du système vasculaire combine une série de tests de laboratoire et en particulier des images d'exploration. Depuis 2004⁵, les organisations médicales et de santé ont approuvé le partage d'informations numérisées du patient sous la forme d'historique clinique électronique basé sur un standard de communication pour l'imagerie appelé DICOM⁶ (*Digital Imaging and Communication in Medicine*) [Vos07]. Ce standard a évolué au cours des dernières années pour prendre en charge non seulement des images médicales, mais également d'autres types d'informations médicales comme des vidéos, des signaux et des rapports structurés. On retrouve par exemple : l'identification du patient, les procédures de l'étude, la date de l'étude, l'identification du médecin, l'identification de l'institution et le diagnostic (ou des observations) [Med12].

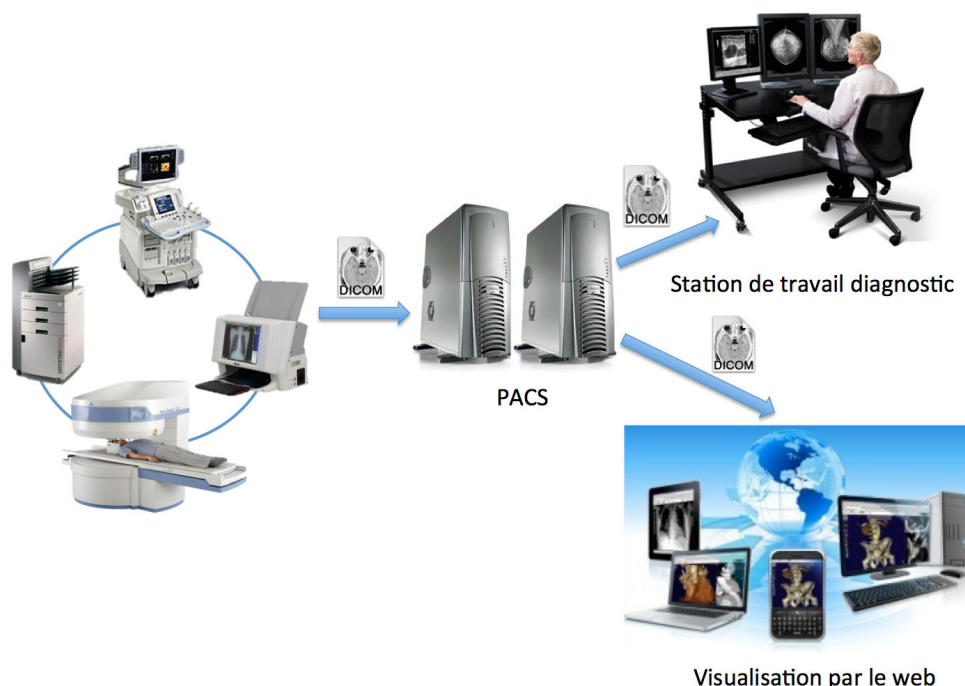


FIGURE 3.11 – Structure générale d'un système d'archivage PACS

Notre objectif est de fournir un scénario de diagnostic évolutif des patients, basé sur DICOM. Ce scénario de diagnostic doit comprendre : l'information de patient, le problème de santé relevant du patient, les examens et les résultats, les diagnostics et leurs images, les traitements et la médication appliquée au patient. C'est donc l'utilisation du potentiel de l'information pour le soutien clinique, la prise de décision et le travail collaboratif.

Avec l'émergence de DICOM, son intégration a été étendue aux dispositifs d'acquisition, de visualisation, d'impression et de stockage des images médicales de différents

5. Réseaux d'imagerie médicale et système d'information au service du patient, <http://fr.calameo.com/read/000162174b7beec2c3258>

6. DICOM, <http://dicom.nema.org/>

fabricants et types d'images (par exemple radiographie numérique (CR), mammographie (MG), Résonance magnétique (IRM), Ultrasons (US), Angiographie digital (XA), médecine nucléaire (NM), ...). Actuellement, les systèmes qui fonctionnent avec des images médicales numérisées peuvent intégrer des données de différents services ou hôpitaux. Le serveur PACS (*Picture Archiving and Communication System*) est un système d'archivage et de transmission d'images lequel permet de gérer les images médicales sur modalités (scanner CT, MRI, PET), d'archiver les images produits, d'utiliser le réseau web pour consulter les images (Figure 3.11).

Par conséquent, il existe un grand nombre d'informations au format DICOM mais l'utilisation de ces productions dans la pratique clinique est généralement limitée au travail individuel des professionnels et de manière non collaborative. Cependant, la recherche médicale nécessite de consolider l'information centralisée par l'extraction de caractéristiques et la validation de techniques et de diagnostics. L'exécution de ces activités est réalisée manuellement et sans outils spécialisés [Sal12].

Le standard DICOM fournit des services pour lier les différents systèmes d'information dans un hôpital, comme par exemple les systèmes d'information de radiologie et les systèmes d'information hospitaliers. La Figure 3.12 illustre le modèle de DICOM, dans lequel un patient a un ou plusieurs dossiers médicaux et chaque dossier contient une ou plusieurs séries d'images. Par exemple, un patient a une relation directe à une ou plusieurs études, mais une étude est associée à un patient unique.

3.3.2.1/ REPRÉSENTATION DU DIAGNOSTIC SUR UN MODÈLE ONTOLOGIQUE

Dans un cheminement de pensée décisionnelle classique utilisé par les médecins, l'étude des symptômes, les analyses et l'observation des résultats qui conduisent au diagnostic sont accompagnés d'une procédure de soin adaptée. Ce cheminement se rapproche du modèle SOAP⁷ (**S**ubjective, **O**bjective, **A**ssise et **P**lan), lequel est utilisé par certains médecins pour décrire l'état de santé complet d'un patient. Dans un tel modèle, la partie Subjective décrit les symptômes et les signes cliniques du patient, la partie Objective décrit les observations et les analyses des médecins, la partie Assise correspond aux maladies et aux problèmes de santé du patient et enfin la partie Plan correspond aux procédures de soins ou au plan de santé mis en place pour résoudre les problèmes du patient.

Une ontologie pour la modélisation du diagnostic doit permettre de représenter un cheminement logique d'une procédure classique de soin. Cette démarche permet, au travers d'un modèle informatique, la représentation d'une prise de décision formalisée pour différents cas et différents types d'imagerie DICOM.

Le flux d'informations et de connaissances qui peut être inclus dans les documents numérisés (DICOM) au sein d'un modèle ontologique permet le partage des données et des connaissances entre les médecins sur un travail collaboratif (Figure 3.13). L'information partagée doit être intégrée dans une ontologie qui considère toutes les situations possibles de diagnostic pour que les données soient effectivement partagées correctement. Cette ontologie du diagnostic doit comprendre : (1) l'information du diagnostic qui inclut la maladie et les analyses d'un patient ; (2) les symptômes qui représentent les

7. Structuration et concepts du dossier médical électronique, <http://www.maisonmedicale.org/Structuration-et-concepts-du.html#nh1>

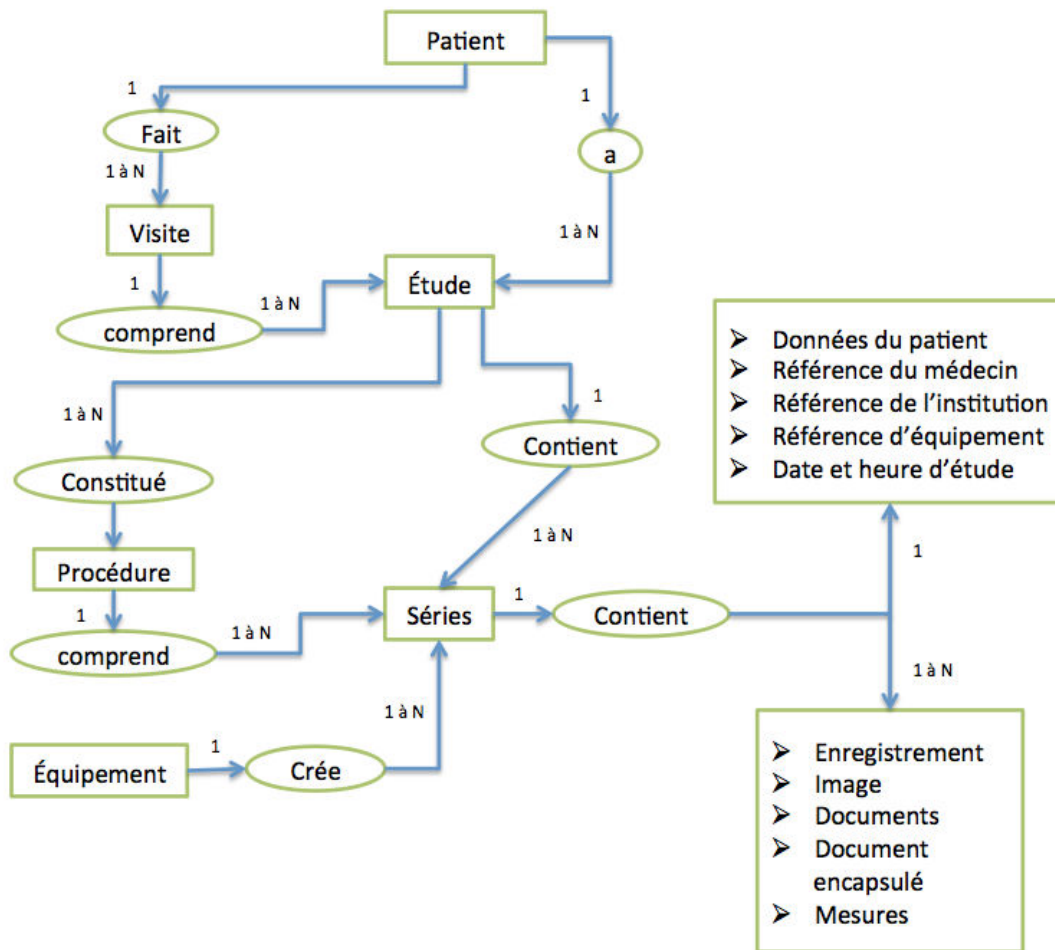


FIGURE 3.12 – Génération d'une étude sur un modèle de DICOM

malaises et les problèmes physiques du patient ; (3) les traitements représentés par les médicaments ou thérapies prescrits pour la maladie ; et (4) les observations ou notes du médecin.

Il existe un problème principal pour la spécification des diagnostics. En effet, un médecin peut définir le diagnostic de différentes manières. Par exemple, le médecin peut utiliser les symptômes comme élément de plus haute importance pour référencer les maladies qui référencent les traitements, tandis qu'un autre peut utiliser la maladie pour organiser les symptômes et leurs traitements. Le problème est la recherche et la sélection des données, car les traitements peuvent être associés à différentes maladies qui ont des symptômes similaires. La complexité d'adaptation de l'information est aussi structurelle. L'information peut avoir différentes hiérarchies au sein de l'ontologie. Pour cette raison, il est très important que l'ontologie du diagnostic soit unifiée syntaxiquement et sémantiquement pour s'adapter aux différents diagnostics.

L'adaptation des éléments de diagnostic à une ontologie OWL basée sur des classes, des attributs et des relations doit représenter également les caractéristiques et les propriétés du diagnostic d'un patient. L'ontologie comprend : l'information du patient, l'information du médecin et l'institution qui contrôle la visite du patient, la maladie traitée, la procédure et le type d'imagerie, les observations et diagnostic, le traitement ainsi que

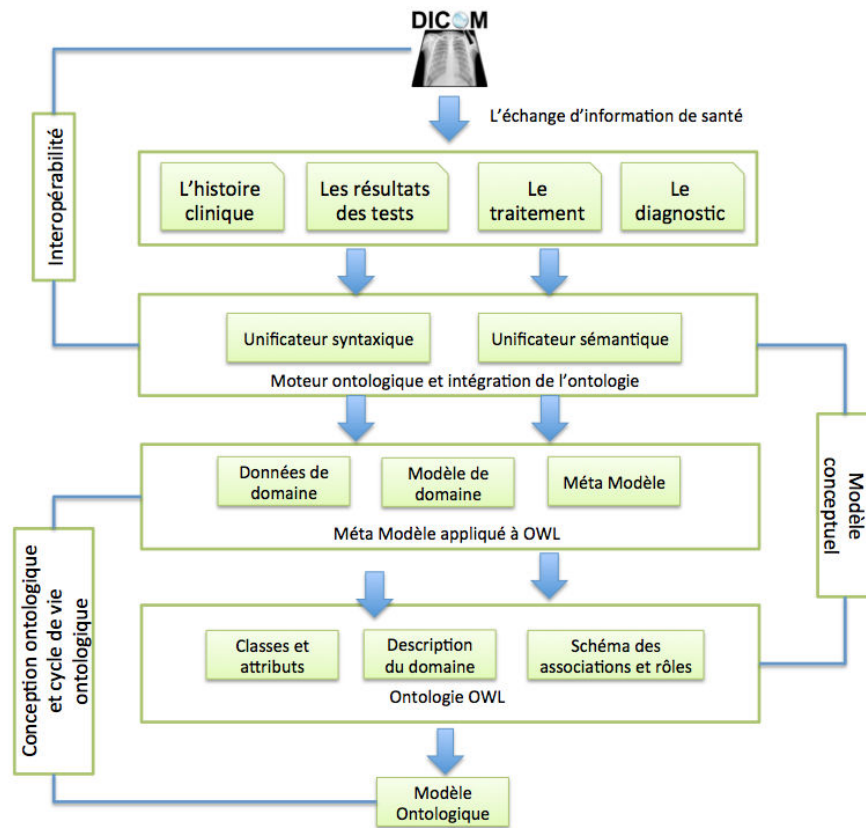


FIGURE 3.13 – Conception des données et flux ontologique

la date d'élaboration du diagnostic. La Figure 3.14 représente un modèle UML pour le groupement des éléments de diagnostic. Les lignes rouges représentent les actions du médecin, et les lignes violettes représentent pour le patient les indicateurs de la maladie ainsi que les thérapies et procédures médicamenteuses mises en place.

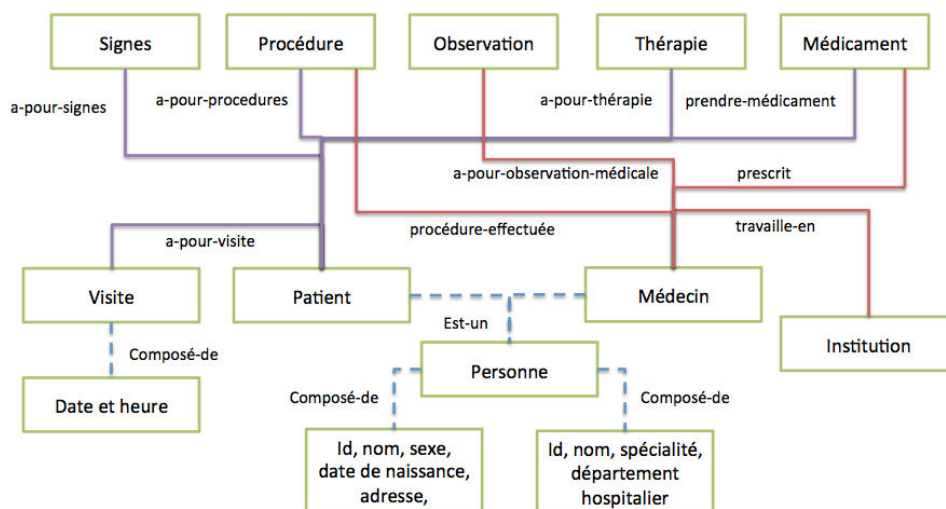


FIGURE 3.14 – Ontologie conceptuelle du diagnostic

La classe *Patient* est définie en utilisant divers attributs (id, nom, adresse, sexe, date de naissance, ...) qui correspondent à des données statiques dans DICOM. La classe est liée à l'association de données du patient et à *Visite* qui contrôle la date du diagnostic et le lien Médecin – Patient.

La classe *Médecin* a les mêmes attributs que *Patient* car les deux sont des personnes associées à un même diagnostic. La différence entre *Patient* et *Médecin* est que le *Patient* est directement lié à une *Visite* et le *Médecin* est lié à une *Institution*. On peut affirmer la relation suivante : « Le Patient (données du patient) a une Visite (date du diagnostic) chez le Médecin (données du médecin) qui travaille dans l'Institution (données de l'hôpital ou clinique où il a effectué le diagnostic) ».

Les classes *Procédure*, *Observation*, *Traitement*, *Thérapie* ont un lien direct avec l'ontologie du système vasculaire pour la représentation exacte des données pour un diagnostic complet (Figure 3.15)

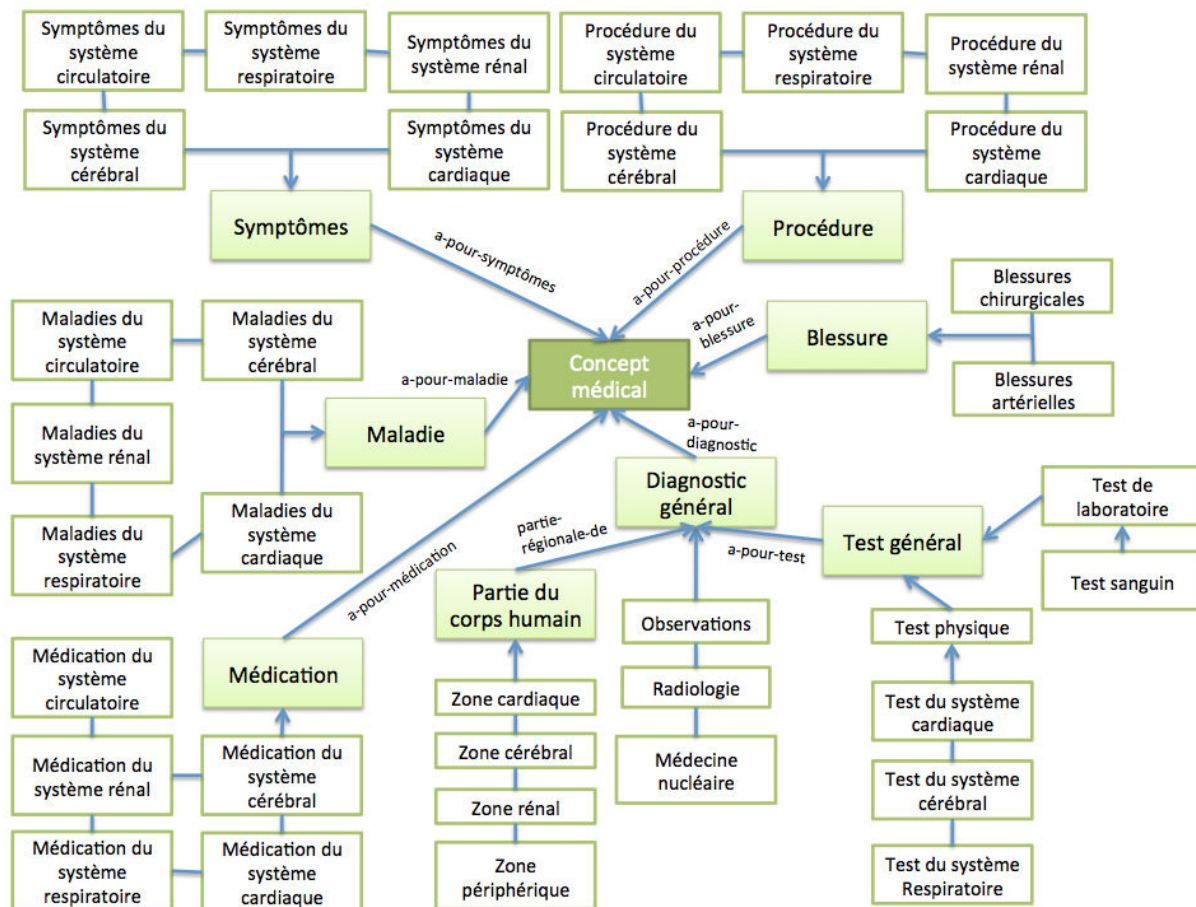


FIGURE 3.15 – Liaisons entre maladies vasculaires et diagnostic

La logique de description sur une procédure *Heart Valve* (présence d'une maladie : Stenosis, Thrombosis, Heart Insufficiency, ...) pour montrer la relation Diagnostic – Procédure – Maladie est la suivante :

$$\text{Heart Valve Operations} \sqsubseteq \text{Heart Procedures} \sqcap$$

$$\begin{aligned}
& \exists Has\ System\ Involvement .\ Circulatory\ System \sqcap \\
& \forall Has\ System\ Involvement .\ Circulatory\ System \sqcap \\
& \exists Localized\ In .\ Heart\ Valve\ Structure \sqcap \\
& \forall Involves\ Dysfunction (Stenosis \sqcup Thrombosis \sqcup Insufficiency) \sqcap \\
& \exists InvolvesAct .\ Replacement \sqcap \\
& \exists InvolvesAct .\ Resection \sqcap \\
& \exists InvolvesAct .\ Repair \sqcap \\
& \exists InvolvesAct .\ Excision \sqcap \\
& \exists InvolvesAct .\ Inspection \sqcap \\
& \forall InvolvesAct (Replacement \sqcup Resection \sqcup Repair \sqcup Excision \sqcup Inspection)
\end{aligned}$$

La représentation basée sur la LD décide que certaines restrictions ajoutent par défaut le constructeur \forall . Dans un diagnostic, il est également nécessaire de savoir quel organe présente une insuffisance.

$$\begin{aligned}
& Organ \sqsubseteq Anatomical\ Structure \sqcap \\
& \exists Regional\ Part\ of (Organ \sqcup OrganSystem) \sqcap \\
& \forall Regional\ Part\ of (Organ \sqcup OrganSystem) \sqcap \\
& \exists Part\ of (Organ \sqcup OrganSystem) \sqcap \\
& \forall Part\ of (Organ \sqcup OrganSystem)
\end{aligned}$$

Finalement, la Figure montre le modèle ontologique du diagnostic qui capture les connaissances nécessaires pour identifier l'état physique d'un patient (à l'aide de l'outil *Protégé-2000*) (Figure 3.16). Cet ontologie contient 91 éléments y compris les classes et sous-classes nécessaires pour contrôler les diagnostics du système vasculaire.

Cette ontologie a été créée pour la modélisation des connaissances et le raisonnement, en s'appuyant la plupart du temps sur l'analyse des textes de *Texas Heart Institut*⁸. Et elle est validée par des experts dans le domaine cardiovasculaire et en médecine interne. Nous avons conçu l'ontologie à partir de descriptions d'expériences médicales en associant les caractéristiques des maladies vasculaires à des méthodes de diagnostic (Figure 3.17).

8. Texas Heart Institut, <http://www.texasheart.org>

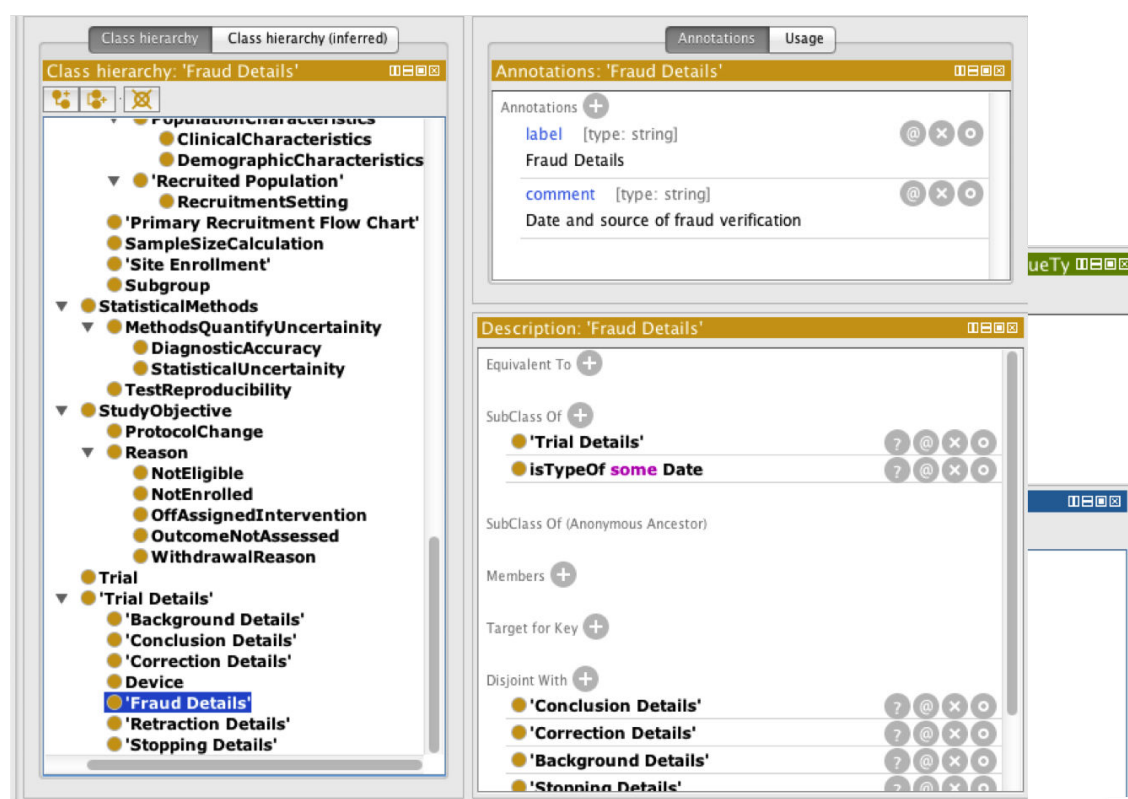


FIGURE 3.16 – Implémentation dans protégé de l'ontologie du diagnostic

3.4/ DÉFINITION D'UNE TROISIÈME ONTOLOGIE DÉDIÉE À LA TRAÇABILITÉ

Un diagnostic médical efficace doit aujourd'hui intégrer des analyses multidisciplinaires. Un système de traçabilité des données du dossier « patient-maladies-traitements » prévoit l'intégration de multiples éléments générés par des disciplines très différentes incluant l'historique clinique du patient (antécédents médicaux et chirurgicaux), l'imagerie médicale (scanner, radiologie, résonance, doppler, ...), les examens de laboratoire (analyse sanguine, histologie, ...) et un processus d'identification du patient/médecin/institution pour un suivi médical complet qui implique la connaissance de l'évolution du patient.

Dans le cadre de l'imagerie médicale, de nombreux progrès médicaux ont été possibles grâce aux évolutions de la transmission des informations, aux évolutions technologiques et au développement des modalités numériques. La numérisation des images et du diagnostic dans un même standard (DICOM) a eu comme effets de réduire les coûts, de contribuer au développement durable, d'optimiser le stockage et surtout a permis d'affiner le diagnostic via les post-traitements des images et la traçabilité de l'état clinique du patient permettant une meilleure cohérence du parcours de soins.

Ce type de système de traçabilité est mis en place pour réduire les erreurs médicales. Le suivi de l'évolution d'un patient sur ses interventions médicales est aujourd'hui possible en grande partie grâce à la télé-médecine et à la collaboration entre médecins.

La télé-médecine permet ainsi au médecin de rester en contact direct avec le patient,

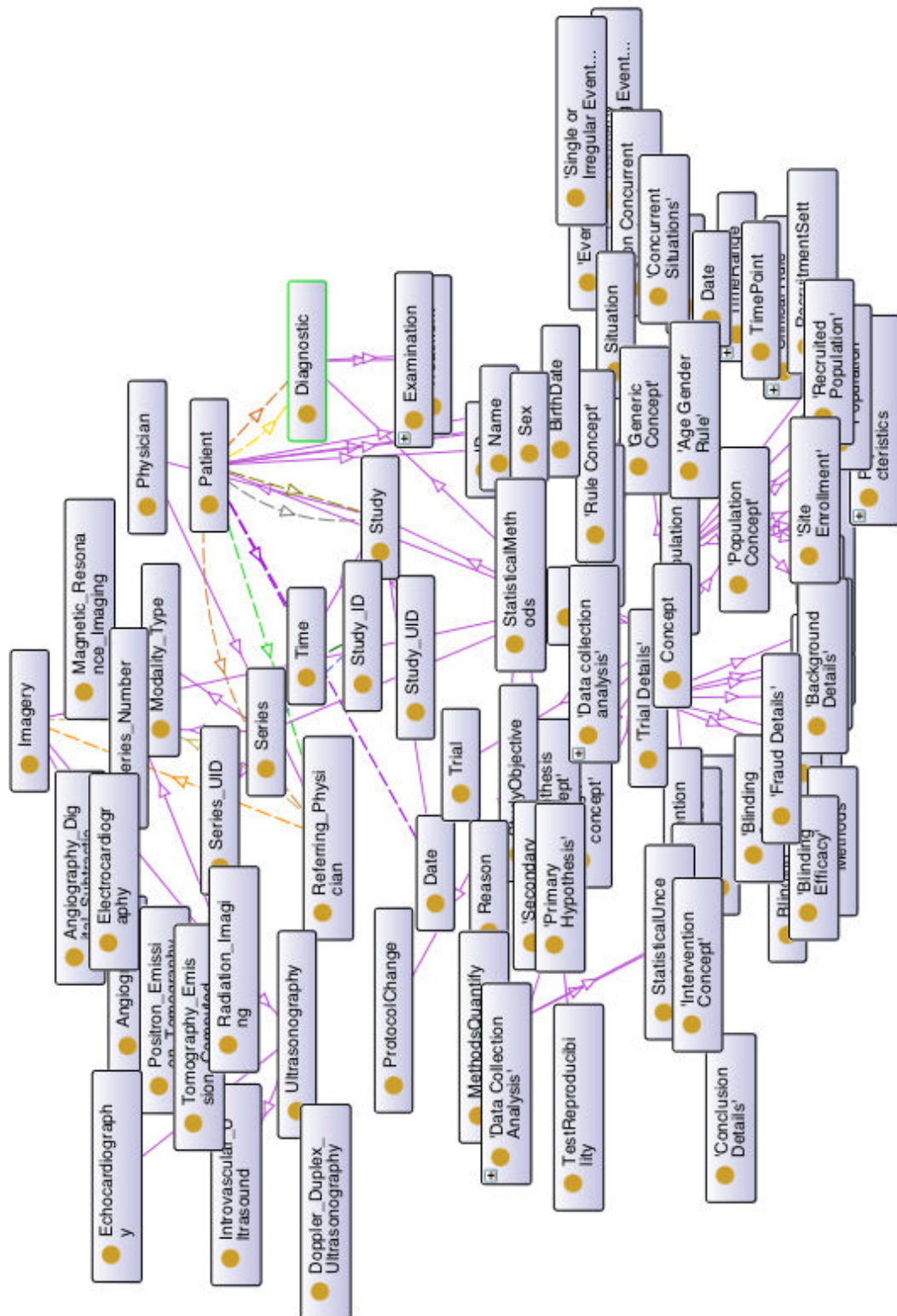


FIGURE 3.17 – Ontologie du diagnostic

mais également de disposer de l'avis d'un autre médecin spécialisé situé à distance du lieu de réalisation du diagnostic. Cette communication a l'avantage de favoriser les échanges de connaissances et de savoir-faire entre les médecins pour la prise de décision du diagnostic et du traitement.

Le recours à la télé-médecine doit s'exercer dans le respect d'un certain nombre de principes pour garantir le bon usage des dossiers des patients et assurer la pratique raisonnée et maîtrisée des transmissions de données sensibles.

L'objectif d'une ontologie de traçabilité du diagnostic est de conserver l'ensemble des actions d'expertises du diagnostic collaboratif, avec un principe d'optimisation de l'accessibilité aux données de grandes tailles qui sont manipulées et échangées. Car, comme nous l'avons montré dans l'état de l'art (cf. 1.3.3, la traçabilité a été identifiée par le Conseil National de l'Ordre des Médecins comme indispensable. Et grâce aux ontologies, l'accès aux traces se fera de manière optimisée.

3.4.1/ TRAÇABILITÉ ET TEMPORALITÉ EN TÉLÉMÉDECINE

La sécurité, la traçabilité et l'efficacité de la télé-médecine apparaissent comme un avantage compétitif pour la surveillance et la certification de l'information transmise. Dans les systèmes médicaux, les tâches les plus importantes sont d'identifier, de classer et de protéger l'information en utilisant des bases de données relationnelles. La traçabilité de l'information médicale a une fonction stratégique qui est le suivi des activités de procédures et des traitements faits par les différents départements hospitaliers et les différents professionnels de santé [Khan13].

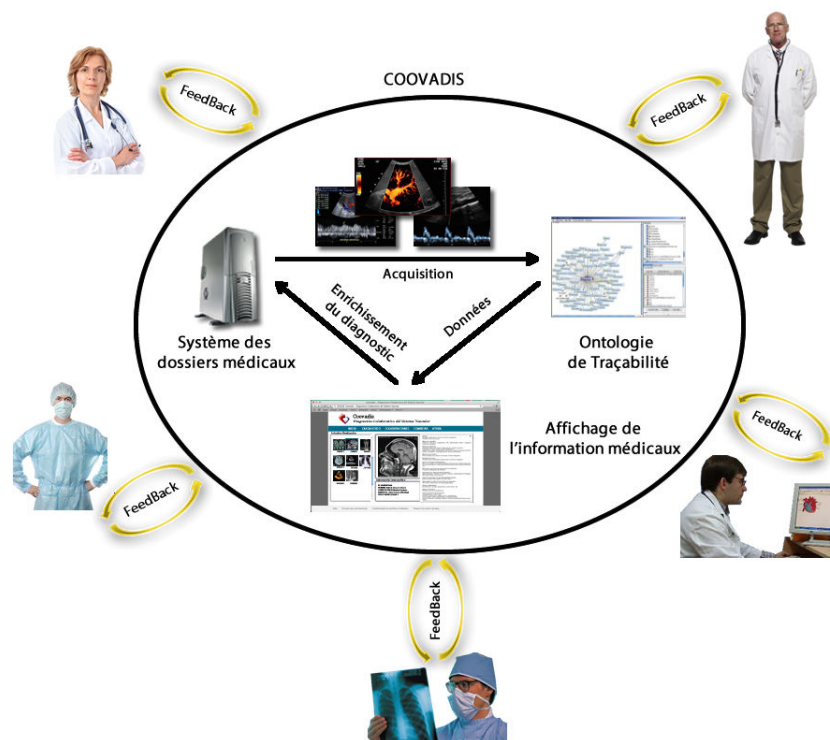


FIGURE 3.18 – Le travail collaboratif sur l'amélioration du diagnostic

L'informatisation des données de santé, en France et Europe, est protégée par un cadre

juridique [OJEU13], qui décrit le processus de traçabilité et définit les conditions minimums qui doivent être garanties, c'est-à-dire : (1) l'authentification des professionnels de santé intervenant dans l'acte ; (2) l'identification du patient ; (3) l'accès des professionnels de santé aux données médicales du patient nécessaires à la réalisation de l'acte.

Le développement d'une ontologie dédiée à la traçabilité du diagnostic et le suivi médical des patients a pour principal objectif : l'amélioration de la qualité des diagnostics dans le domaine vasculaire. L'ontologie de traçabilité avec l'ontologie du diagnostic va permettre un post-traitement du diagnostic dans un travail collaboratif sous la responsabilité des médecins des différentes spécialités. L'application de cette ontologie doit s'intégrer dans des environnements informatiques qui permettent une meilleure interprétation des résultats de l'imagerie et du diagnostic associé (*cf.* Chapitre 4).

3.4.2/ ONTOLOGIE DE TRAÇABILITÉ

L'objectif principal de cette ontologie est de permettre un échange d'information entre les professionnels de santé qui contribuent à l'amélioration du diagnostic (Figure 3.18) par le partage d'expériences et principalement l'aide à la prise de décision clinique. Pour obtenir une collaboration clinique, il est nécessaire de pouvoir : (1) Récupérer les différents diagnostics liés au patient tout au long de sa gestion clinique ; (2) Récupérer les traitements et procédures liés au diagnostic ; (3) Afficher la chronologie de participation des médecins qui ont contribué aux diagnostics ainsi que leur institution hospitalière et (4) Afficher les commentaires, les observations ou les recommandations cliniques que les médecins ont proposés.

L'enjeu principal de cette ontologie constitue un facteur clé d'amélioration de la performance du diagnostic. Son utilisation dans le travail collaboratif est la réponse organisationnelle et l'unification entre Diagnostic–Collaboration (Ontologie de diagnostic – Ontologie de traçabilité). Les objectifs d'utilisation de cette ontologie sont :

- l'amélioration de l'accessibilité de tous les médecins travaillant de manière collaborative,
- l'impulsion d'une meilleure coordination entre les médecins (enregistrement de : actions de participation, contenus, informations du spécialiste, dates et heures de participations),
- la favorisation au recours maîtrisé du diagnostic du système vasculaire numérisé,
- La prise en compte des besoins et attentes du patient,
- L'amélioration en toute sécurité du partage de l'information entre les professionnels de santé,
- La favorisation d'une meilleure prise de décision clinique.

La Figure 3.19 montre une représentation de l'information dans l'ontologie de traçabilité. Cette ontologie fournit l'information nécessaire pour la distribution du diagnostic et la modélisation du travail collaboratif entre les professionnels. L'ontologie de traçabilité est liée directement à l'ontologie du diagnostic et a quatre fonctions principales : (1) Décrire les différents types de dysfonctionnement dans le système vasculaire ; (2) Donner un aperçu de tous les examens et traitements associés au patient ; (3) Montrer les institutions et les médecins qui ont collaboré pour améliorer le diagnostic et les observations individuelles pour la prise de décision clinique ; et (4) Montrer la participation (en enregistrant la date) au diagnostic.

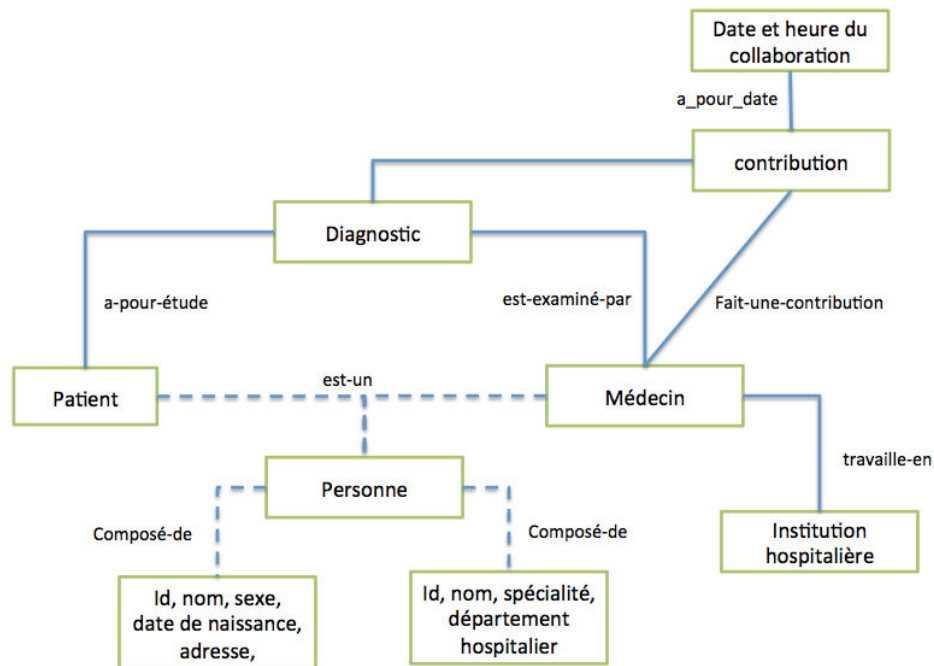


FIGURE 3.19 – Diagramme conceptuel de l'ontologie de traçabilité

L'intégration de différentes sources d'information dans le cadre médical en utilisant des ontologies (Maladie–Diagnostic–Traçabilité) est rendue possible grâce à la communication entre les ontologies (Figure 3.20), c'est-à-dire l'information est récupérée et organisée grâce à la communication entre les ontologies permettant l'analyse de différents facteurs, vocabulaires et rôles. Cette information est présentée de manière à faciliter la compréhension de l'utilisateur (professionnels médicaux).

L'ontologie de traçabilité est constituée des composants suivants (Figure 3.21) :

Dans le contexte de l'aide à la décision clinique, cette ontologie permet d'accéder à l'information et d'aider la communication entre les médecins. Elle fournit des connaissances pertinentes sur les soins à donner aux patients. Elle a l'avantage de combiner les informations relatives au diagnostic, de faciliter l'accès aux données du patient et d'identifier les indicateurs dans les données du patient pour améliorer le diagnostic. Principalement, cette ontologie fournit une terminologie standard pour les concepts médicaux qui peuvent faciliter l'intégration de données pour le travail collaboratif.

L'ontologie complète de traçabilité (Figure 3.22 et Figure 3.23) correspond à la modélisation de la collaboration du diagnostic entre médecins.

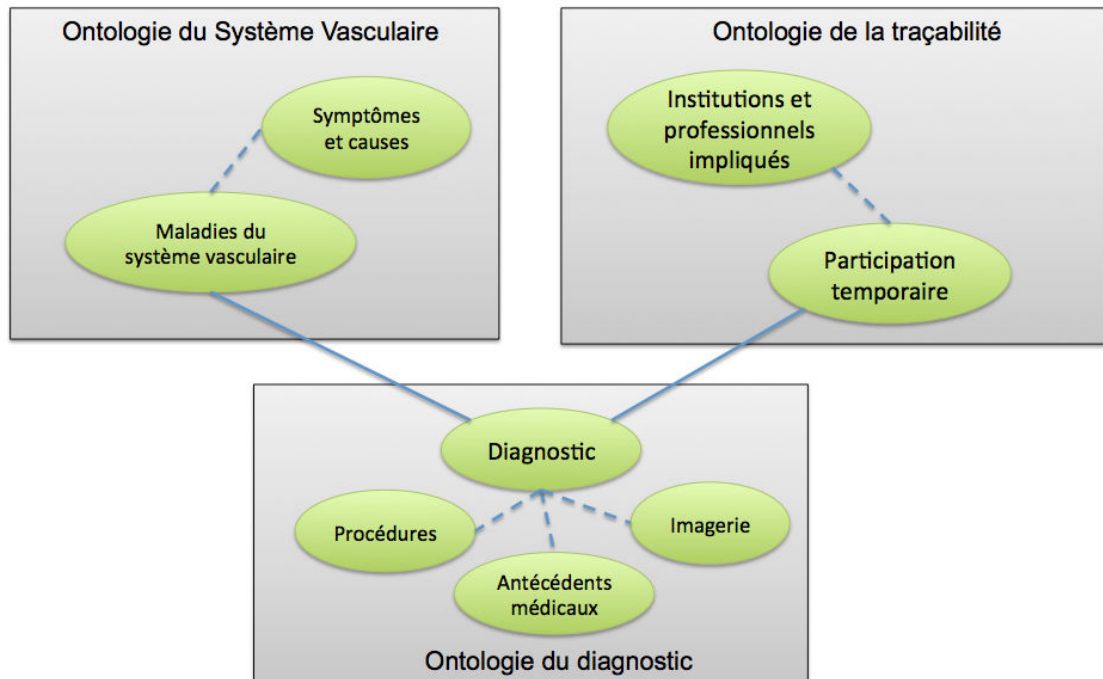


FIGURE 3.20 – Diagramme conceptuel de l'ontologie de traçabilité

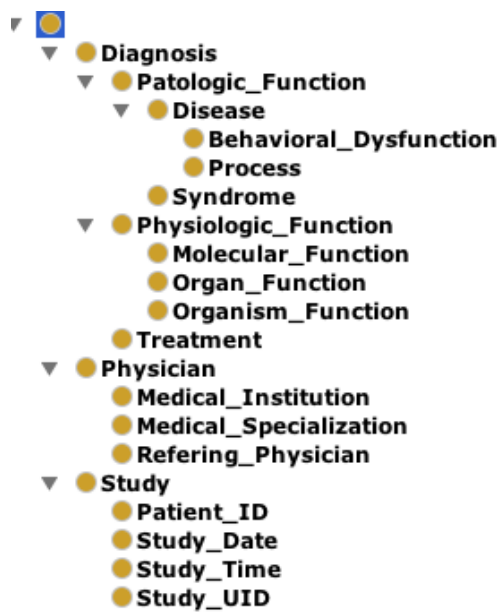


FIGURE 3.21 – Composants de l'ontologie de traçabilité

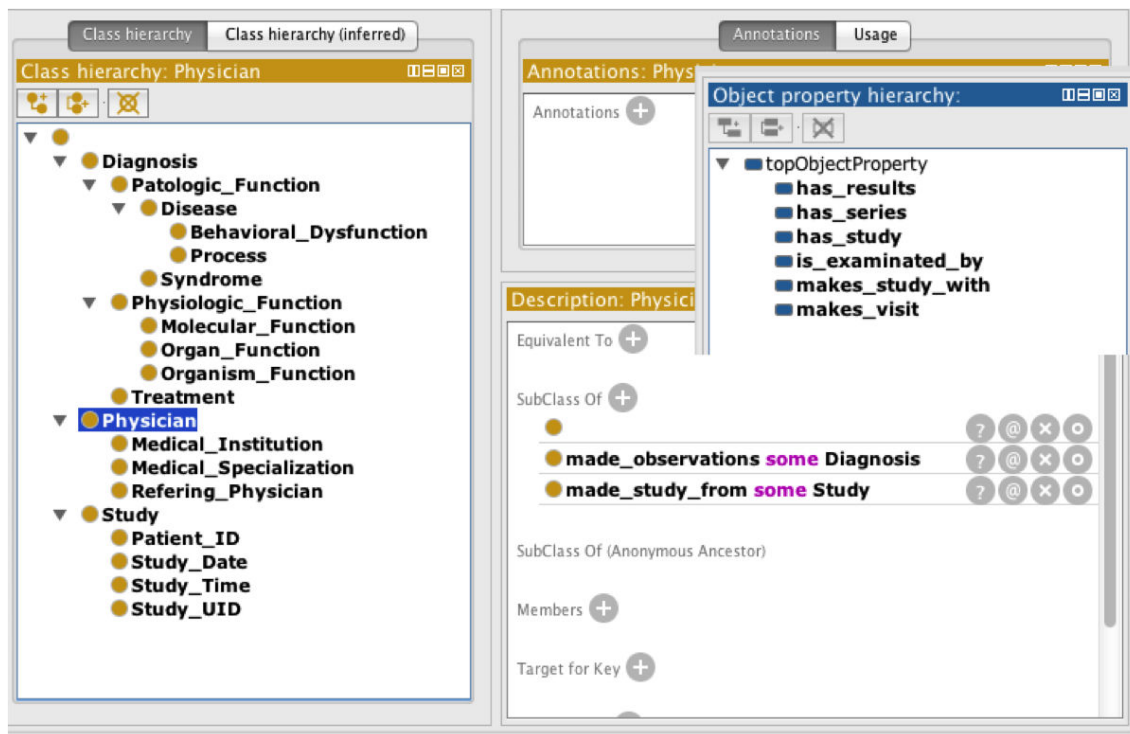


FIGURE 3.22 – Implémentation dans protégé de l'ontologie de traçabilité

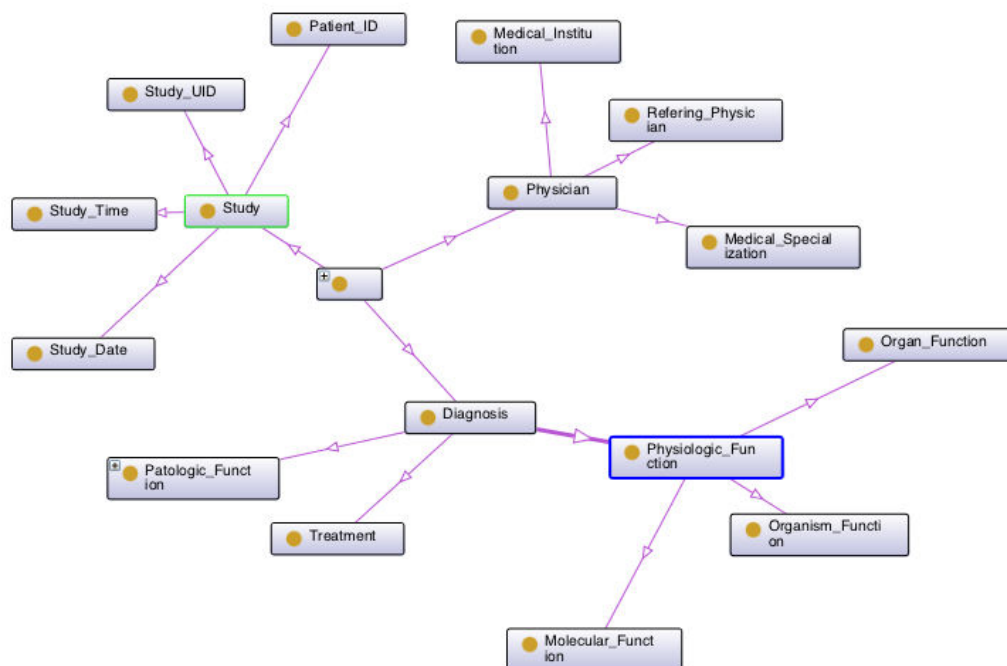


FIGURE 3.23 – Ontologie de traçabilité

CONCLUSION



Nous avons présenté, dans ce chapitre, la démarche que nous avons utilisée pour le développement de nos trois ontologies : l'ontologie du système vasculaire, l'ontologie du diagnostic et l'ontologie de traçabilité.

Nous avons adopté une démarche méthodologique, dites *à partir de rien* pour leur développement. Ces ontologies nous permettent d'effectuer la caractérisation sémantique d'informations médicales pour l'aide à la collaboration des diagnostics et décisions cliniques. L'ontologie de diagnostic et l'ontologie de traçabilité ont pour but l'indexation des connaissances à travers une langue de spécialité (vocabulaire, terminologie et sémantique spécifiques du diagnostic) pour le travail collaboratif.

Le chapitre suivant est consacré à notre nouvelle plateforme *COOVADIS : COllabOrative VAscular DlagnoSis*

COOVADIS : *COllabOrative VAScular DlagnoSis*

INTRODUCTION

Ce chapitre est consacré à la définition de COOVADIS, notre plateforme d'aide à la collaboration entre professionnels de santé pour l'interprétation des données, l'amélioration et la traçabilité du diagnostic d'un patient. Cette plateforme développée en mode SaaS (*Software as a Service*) utilise comme principale ressource les ontologies du système vasculaire, de diagnostic et de traçabilité que nous avons décrites dans le chapitre précédent. Cette plateforme est accessible par un navigateur web. Elle permet la connexion à distance depuis différents dispositifs (ordinateur, tablette ou téléphone tactile, ...) : et ainsi le diagnostic collaboratif à distance *tracé* est favorisé, les praticiens pouvant se connecter en toute mobilité à la plateforme pour travailler ensemble de manière collaborative.

nous avons défini dans le chapitre précédant les ontologies qui constituent le cœur fonctionnel de notre plateforme, à l'aide de ces descriptions préalables nous présentons maintenant la plateforme dans sa globalité.

Nous présentons dans ce chapitre quatre sections : une première pour définir la plateforme ainsi que son architecture et une deuxième, plus technique, qui expose les outils et la manière dont nous avons implémenté notre plateforme, la troisième qui présente la plateforme Web finale et ses fonctionnalités, et enfin la quatrième qui présente un cas clinique (notre *fil rouge* Mr J.).

4.1/ DÉFINITION DE LA PLATEFORME

Faciliter les interactions entre médecins peut être une réponse aux besoins et aux exigences du système de santé d'aujourd'hui. En particulier, ce type d'interaction aborde la nécessité d'un accès à l'information médicale sans restriction aux services de santé et facilite la gestion du flux de travail dans une plateforme médicale. L'objectif est de favoriser les interactions entre médecins par un mécanisme solide (cloud computing basé sur un mode SaaS) et un moteur de recherche ontologique des connaissances cohérentes qui permet de créer de nouvelles connaissances pour le traitement des maladies complexes. La plateforme exploite le standard DICOM qui est une représentation structurée des examens du patient, d'imagerie et du diagnostic liés entre eux.

Dans le cadre de recherches visant à contribuer au partage de connaissances entre médecins, notre objectif est de proposer une solution de communication entre médecins et d'indexation sémantique des examens médicaux du patient, standardisée selon la norme DICOM. Le recours aux ontologies est essentiel pour répondre aux besoins des professionnels de la santé, dans un but d'aide au diagnostic et de travail collaboratif.

Cette plateforme dispose de quatre modules (Figure 4.1) : ❶ un module pour l'administration, ❷ un module pour l'indexation sémantique, ❸ un module pour la recherche des patients et de leurs diagnostics associés et ❹ un module pour le travail collaboratif entre professionnelles de la santé pour l'amélioration des diagnostics.

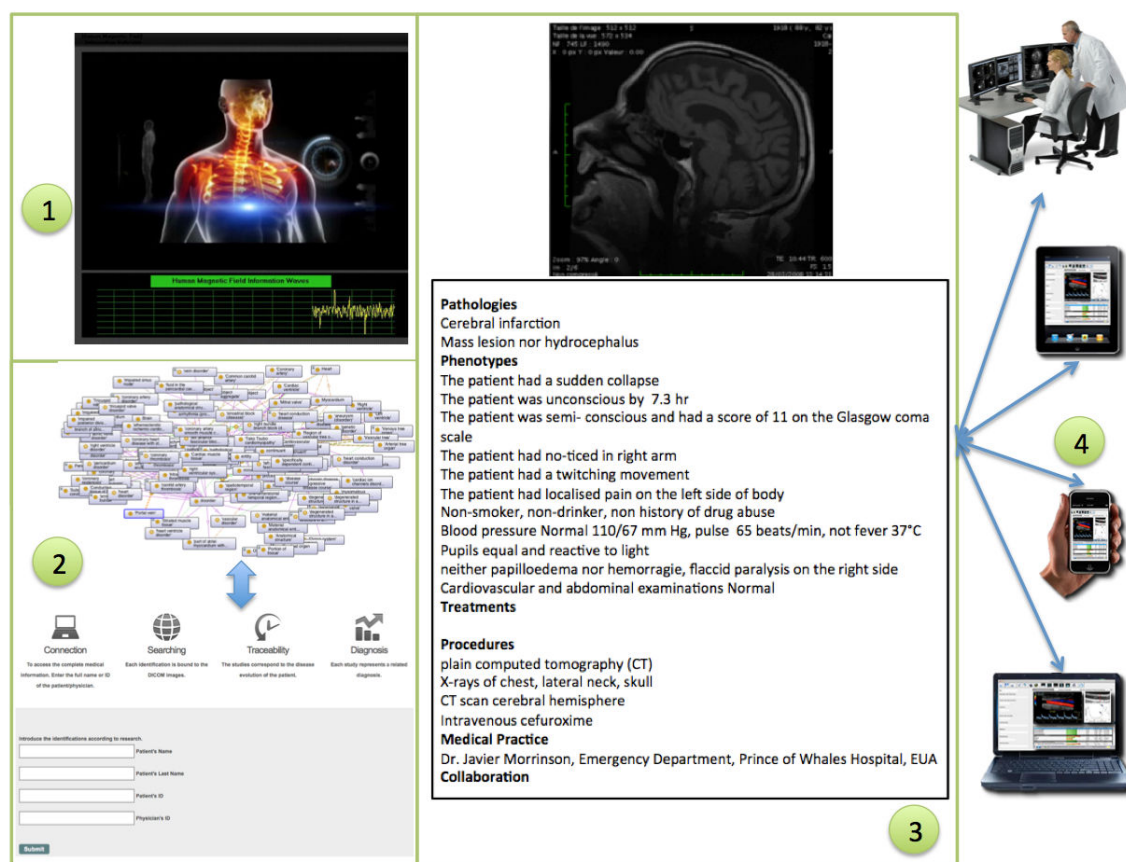


FIGURE 4.1 – Architecture globale de COOVADIS

4.1.1/ COOVADIS : UNE AIDE À LA COLLABORATION DU DIAGNOSTIC

COOVADIS (**CO**llab**OR**ative **VA**scular tele**DI**agno**SIS**) est une plateforme dédiée au diagnostic collaboratif de maladies du système vasculaire. Cet environnement combine les avantages de l'ingénierie ontologique et du web sémantique. Elle permet à la communauté médicale l'accès aux diagnostics, l'aide à la collaboration entre médecins pour l'élaboration d'un diagnostic (accès sécurisé). La plateforme interprète la requête de recherche de patient, catégorise les diagnostics liés au patient et affiche les évolutions de la pathologie du patient. La catégorisation et l'indexation de la connaissance est réalisée par la manipulation des ontologies (ontologie du système vasculaire, ontologie du diag-

nostic et ontologie de traçabilité, cf. Chapitre 3) en utilisant des images au format DICOM. Pour aider à la réalisation d'un diagnostic, COOVADIS a comme objectif la corrélation de l'information des diagnostics et la création dynamique de connaissances au cours des interventions temporaires, c'est-à-dire, des connaissances créées durant un travail collaboratif où chaque spécialiste peut partager des informations, des expériences et des expertises pour la prise de décisions cliniques. Ainsi la plateforme permet aux professionnels impliqués dans l'aide au diagnostic collaboratif d'avoir un accès total aux rapports médicaux mis à jour et aux images exploratoires DICOM associées.

Comme nous le montre l'état de l'art du travail collaboratif (cf. Section 1.1) les points clés du travail collaboratif sont le partage de l'information, et les échanges cohérents des partages (par exemple des données patient) et des actions (par exemple participation au diagnostic de chaque acteur professionnel de santé).

La Figure 4.2 montre notre objectif principal qui est de mettre en œuvre le partage de l'information entre professionnels de la santé et l'amélioration des diagnostics numériques en appui sur des images DICOM. Cette chaîne de traçabilité contient la catégorisation : de l'information du diagnostic, de l'information de la pratique clinique, des mentions et observations du médecin, des institutions médicales, du type des images d'observation, du fabricant d'appareillage pour l'imagerie médicale et de l'amélioration du diagnostic pour le travail collaboratif.

La valeur ajoutée de cette plateforme est un gain de temps pour la prise de décision clinique sur des diagnostics complexes. La plateforme aide les médecins à :

- partager des informations cruciales,
- faire une interprétation sémantique de l'information,
- faire un rapport automatique pour aider les médecins plus expérimentés,
- obtenir une collaboration médicale sans que le groupe de travail soit physiquement présent à chaque instant.

Cette chaîne de traçabilité permet l'inclusion de tous les diagnostics passés et présents liés au patient pour avoir une meilleure visibilité sur l'évolution de la pathologie grâce au support des ontologies.

Comme nous l'avons montré dans l'état de l'art (cf. Section 1.3.3), la traçabilité a été identifiée par le Conseil National de l'Ordre des Médecins comme indispensable. La réduction des inégalités de ressources entre les établissements de santé figure dans la réforme de l'hospitalisation. La normalisation des diagnostics et des actes médicaux est devenu une obligation légale des structures de soins (cf. section 3.3.2.1).

Notre volonté est que pour proposer un environnement d'aide au diagnostic collaboratif, il faut :

- non seulement permettre aux médecins de se ré-appropriier le diagnostic du patient en mettant en place un langage médical habituel et complet,
- mais également proposer une plateforme semi-automatique pour assister le médecin dans sa tâche de collaboration, d'interprétation du diagnostic et de prise de décision.

Cette plateforme propose aux médecins spécialisés en cardiologie, neurologie, chirurgie vasculaire, médecine interne, urgence... de partager leurs questions, leurs observations et leurs échanges d'avis avec leurs confrères, de manière asynchrone ou en temps réel. Ils peuvent notamment échanger sur les cas cliniques qui leur posent problème, voire même confronter leurs pratiques professionnelles à celles de leurs confrères.

Selon leurs besoins, les médecins peuvent ainsi solliciter l'avis et la collaboration au diag-

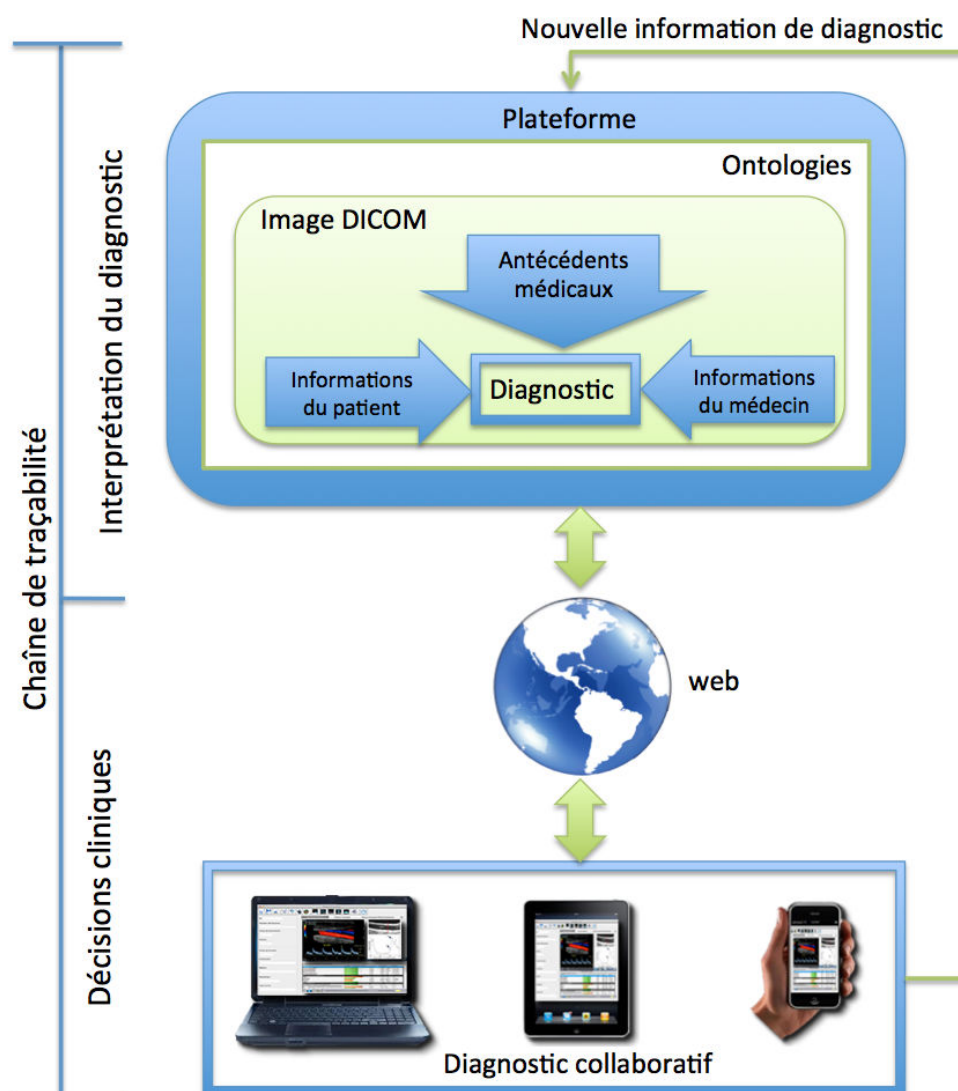


FIGURE 4.2 – Partage de l'information entre professionnels pour l'amélioration des diagnostics

nostic d'un professionnel ou d'une communauté distante, de médecins d'une spécialité, de quelques médecins d'un groupe fermé. Par exemple, un cardiologue pourra simultanément partager l'information d'un traitement avec ses proches confrères au sein d'un même hôpital mais de différents services. Pour cela, un outil dédié à la collaboration pour l'aide au diagnostic, et accessible par internet, permet d'échanger des observations et des images de manière asynchrone afin de répondre aux contraintes de temps des médecins qui ne peuvent que rarement se libérer au même moment que leurs confrères pour s'entretenir ou travailler avec eux.

Dans le contexte du domaine médical, il est important de souligner qu'il existe une grande quantité de termes médicaux qui doivent être gérés. Leur manipulation peut entraîner des problèmes d'interprétation lorsque plusieurs mots sont utilisés pour décrire le même concept (par exemple un examen de radiologie qui spécifie une image par « résonance magnétique », les synonymes en langue anglais sont : MR Tomography, MRI, Zeugmatography, Imaging Chemical Shift... ; Les synonymes de l'« infarctus » en langue anglais

sont : stroke, acute accident, apoplexy, heart attack, infarct, coronary accident, vascular accident. . .). Pour résoudre ce problème, cette plateforme permet l'utilisation d'ontologies qui gèrent la terminologie médicale complète utilisée aujourd'hui par les centres médicaux (ontologies en langue espagnol pour l'évaluation des cardiologues mexicains de Morelia, et ontologies en langue anglaise afin d'être également utilisables en Europe).

4.1.2/ ARCHITECTURE DE LA PLATEFORME

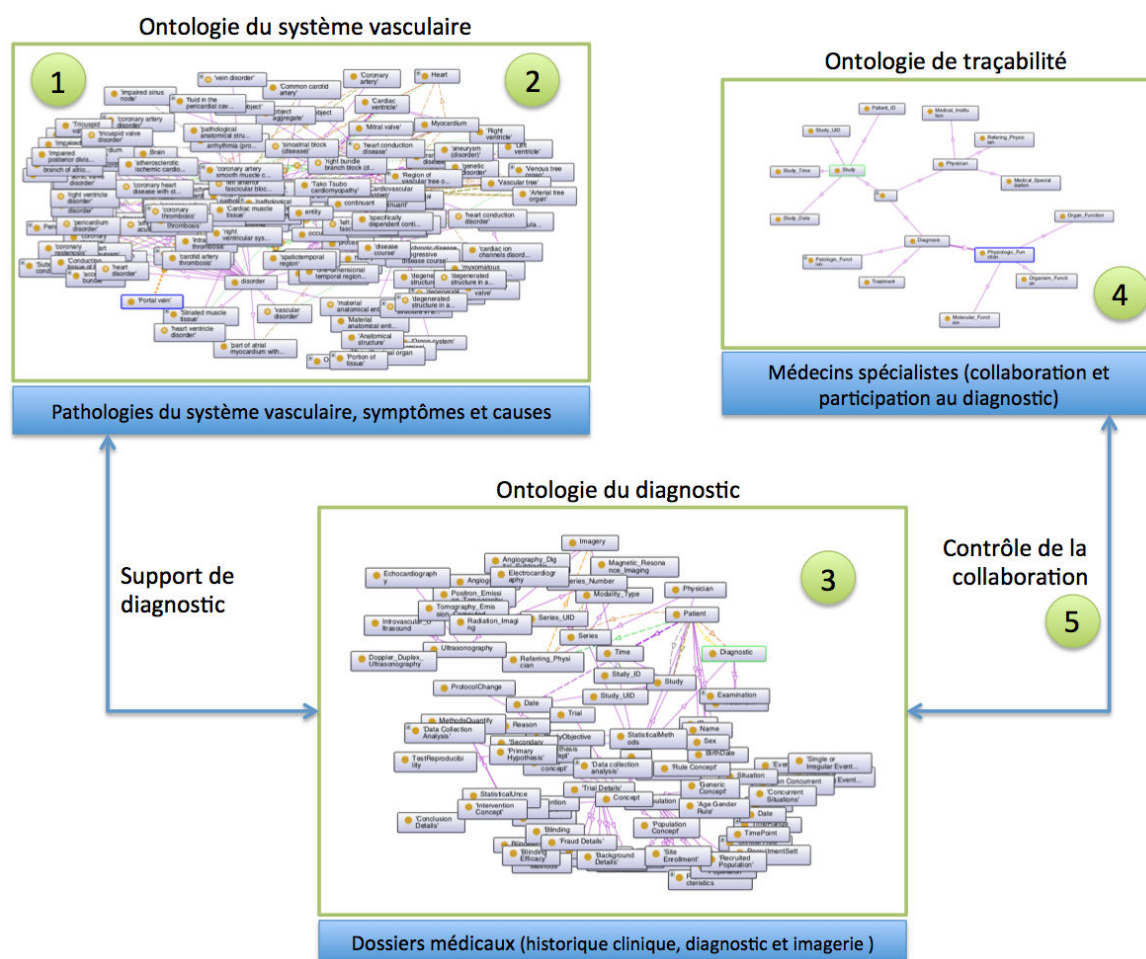


FIGURE 4.3 – Connexion des ontologies dans le plateforme COOVADIS

Il est essentiel pour une collaboration entre médecins distants d'avoir la possibilité d'accéder à l'application par un navigateur Web : en effet, on peut ainsi s'affranchir de l'hétérogénéité des plateformes, les outils personnels de navigation Web se chargeant pour une grande partie de la gérer. COOVADIS est développé en mode *SAAS Software as a Service*, lequel permet un meilleur contrôle de l'information accessible pour la collaboration. Nous avons utilisé également des technologies sémantiques pour le stockage, la représentation et le traitement des informations des diagnostics.

Nous ne souhaitons pas automatiser la procédure de diagnostic. Notre idée est de proposer une représentation des connaissances avec laquelle le médecin peut interagir pour l'aider à construire une représentation de la pathologie du patient, en tenant compte

de ses propres capacités de choix et de l'interprétation du diagnostic. Le problème de l'interprétation informatique se situe au niveau du formalisme et de l'expression des connaissances qui sont présentes le plus souvent en langage naturel sous forme textuelle (termes de représentation du sens). Nous devons donc proposer une plateforme qui permet de réaliser des diagnostics évolutifs au cours du temps et qui permet également aux médecins d'interagir avec le système formel dans des termes de leur domaine de spécialité, répertoriés au sein d'ontologies.

Dans notre recherche, comme nous l'avons exposé au chapitre précédent, nous proposons trois ontologies OWL : ontologie du système vasculaire, ontologie du diagnostic et ontologie de traçabilité. Ces ontologies fournissent les informations nécessaires, distribuées dans les images DICOM, pour permettre un meilleur suivi des patients. Il y a cinq fonctions principales (cf. Figure 4.3) :

- ❶ décrire les différents types de dysfonctionnement du système vasculaire,
- ❷ décrire les symptômes et les causes responsables de la maladie,
- ❸ donner une vue d'ensemble de tous les diagnostics du patient, ainsi que les traitements associés,
- ❹ afficher les institutions et professionnels dans la collaboration pour l'aide au diagnostic,
- et ❺ associer la participation des médecins à un diagnostic.

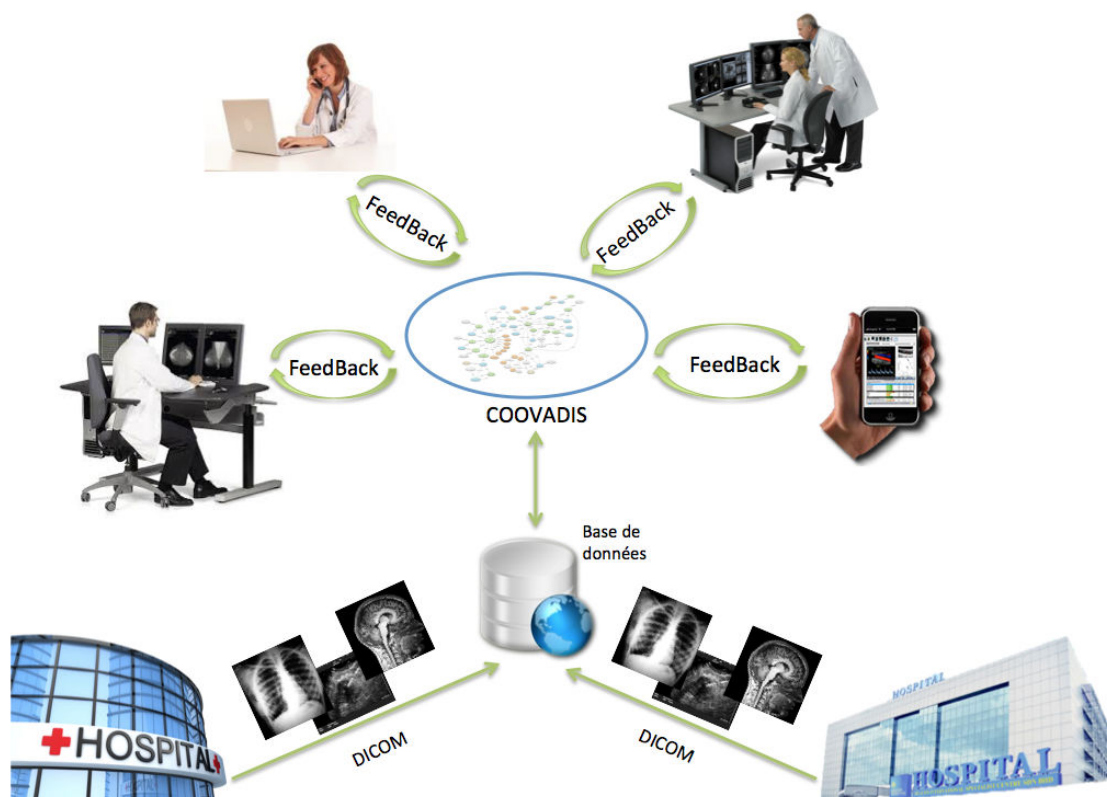
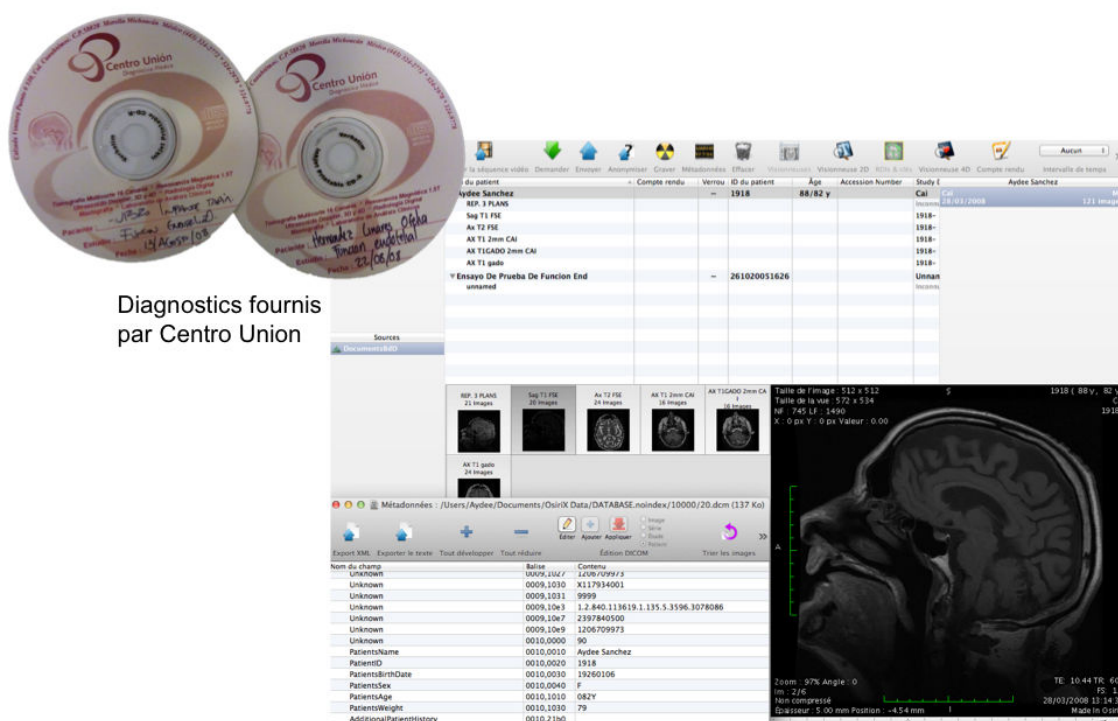


FIGURE 4.4 – Travail collaboratif pour l'ajout de connaissance sur des cas cliniques

Pour obtenir des résultats attendus par la collaboration, la plateforme COOVADIS fournit à chaque médecin au sein de l'environnement de collaboration, la même base de connaissances et elle contient une section pour l'ajout de commentaires et/ou d'observations à

la prise de décisions. Une fois que les médecins ont participé, leurs ajouts seront analysés afin de trouver des similitudes au sein de la base de connaissances ontologiques, qui est considérée comme la base commune de connaissances. Les résultats des nouvelles consultations, sur des mêmes cas cliniques, utilisent cette base de connaissances commune et les résultats s'affichent avec l'ajout de la connaissance avec l'information de chaque participation (Figure 4.4).

Nous disposons également d'un thésaurus spécialisé dans les pathologies du système vasculaire, des diagnostics numérisés sur le standard DICOM (fournis en langue espagnol) mis au point et fourni par le *Centro Unión-diagnostico médico* et par l'*UPAC (Unidad de prevención y atención cardiometabolica)*, institutions de santé de Morelia au Mexique. Ce thésaurus prend en compte les cas de codages médecin-diagnosticien les plus couramment utilisés par les cardiologues et neurologues. Les codages comprennent deux parties : (1) le code de diagnostic selon le codage médecin-radiologiste (Centro Unión) et (2) le code de diagnostic selon les observations des spécialistes traitant des patients (UPAC). Les images DICOM nous servent à enrichir le vocabulaire fourni par les ontologies et par conséquent à améliorer la reconnaissance automatique des expressions pertinentes dans des dossiers médicaux réels. Ces ressources médicales sont montrées dans la Figure 4.5. Ces données (images DICOM) sont stockées dans une base de données MySQL qui contient donc plusieurs examens de patients et leurs descriptions pathologiques.



Métadonnées incluses dans les images DICOM

FIGURE 4.5 – Diagnostics numérisés sur les standards DICOM

4.2/ IMPLÉMENTATION

Dans cette deuxième section du chapitre 4, nous décrivons les choix et outils qui nous ont permis de développer le premier prototype de COOVADIS. Cette section demeure très technique, mais il était important qu'elle figure au sein de ce document pour justifier des choix qui ont mené au produit final qui est utilisé dans le chapitre 5 pour la validation.

Pour la réalisation de notre plateforme nous avons étudié et utilisé différents outils. Ces outils permettent le développement de nos ontologies, mais également la communication entre elles, la réalisation de requêtes, et la représentation des résultats au sein d'une page web. Ils sont présentés ci-dessous :

- Protégé-2000 est un logiciel *open-source* pour l'édition et le développement des ontologies (RDF, RDFS, OWL). Cet éditeur permet de développer une ontologie pour un domaine donné, de définir des formulaires d'entrée de données et d'acquérir des données à l'aide de ces formulaires sous forme d'instances de cette ontologie. Il existe également une librairie Java qui peut être étendue pour créer de véritables applications à base de connaissances en utilisant un moteur d'inférence pour raisonner et déduire de nouveaux faits par l'application de règles d'inférences aux instances de l'ontologie.
- Java, qui par ailleurs devient un standard, est également une plateforme *open-source* qui permet de manipuler de nombreux langages pour les ontologies (Jena, SPARQL), de réaliser le design et la communication web (HTML5, JavaScript, JSP) et de gérer les accès à des bases de données (SQL).
- Jena est une API *open-source* (licence Apache) développée par *HP Labs semantic Web programme*, permettant de lire et de manipuler des ontologies décrites en RDF, RDFS ou OWL, de stocker en mémoire les connaissances et d'appliquer certains mécanismes d'inférences.
- SPARQL est un langage de requêtes orienté données. Il permet d'interroger uniquement les informations détenues dans des modèles ontologiques. Il n'y a pas d'inférence dans le langage de requête lui-même. Le modèle Jena peut être « intelligent » en ce sens qu'il fournit la possibilité de créer certains triplets existant à la demande, y compris le raisonnement OWL. SPARQL prend la description de ce que l'application souhaite, sous la forme d'une requête, et retourne cette information, sous la forme d'un ensemble de liaisons ou d'un graphe RDF.
- Html5-css3-javascript sont des langages qui permettent de décrire la façon dont un document doit être présenté ainsi que les liens qu'il établit. Ils servent à décrire la structure et le style du contenu d'un document et ils communiquent avec les navigateurs pour donner à l'utilisateur final un résultat compressible. Html5 et css3 sont des langages puissants qui permettent de créer des interfaces multi-échelle (échelle de portable, ordinateur, tablette, ...).
- JSP est un langage de programmation qui sert à créer des programmes sur le web. Il permet l'intégration de contenu dynamique. JSP (Java Servlet) permet de programmer simplement l'affichage de contenus dynamiques sur HTML, qui inclut du code Java qui s'exécutera soit sur le serveur Web, soit sur le serveur d'application. Le langage HTML décrit la manière dont s'affiche la page. Le code Java sert, quant à lui, à effectuer

un traitement, par exemple récupérer les informations nécessaires pour effectuer une requête dans une base de connaissances.

- dcm4che est une application *open-source* créée en Java pour l'application du service de santé. C'est une implémentation robuste de la norme DICOM qui permet la gestion de l'image et de son contenu.

L'implémentation de COOVADIS consiste premièrement à l'édition de nos ontologies avec le langage OWL, suivie par son exploitation dans une application de gestion *Jena* et d'annotation *JSP*, afin d'intégrer une application développée pour le travail collaboratif accessible par le web. Les ontologies ont été développées avec le logiciel *Protégé* et la manipulation des ontologies est faite par *Jena 2.6.4* (version la plus récente de *Jena* au moment de l'implémentation). Nous avons également utilisé *JSP* pour les communications réseaux par lesquelles les requêtes *SPARQL* et les réponses sont émises. Il décode également les requêtes faites pour les utilisateurs (médecins) et formate les réponses *SPARQL* au format approprié supporté par le protocole HTTP. Les dossiers des patients intégrés aux images DICOM sont stockés dans une base de données MySQL manipulée par *JSP* grâce à *dcm4che3*. La Figure 4.6 montre l'interaction décrite.

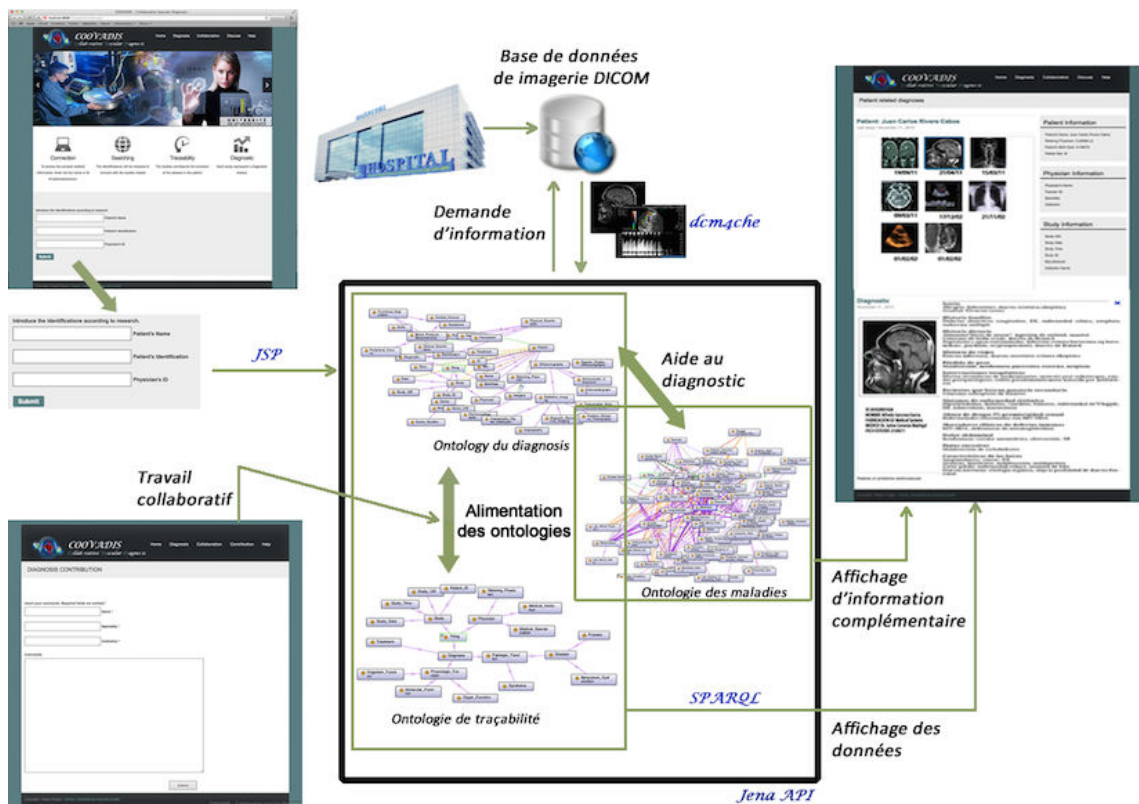
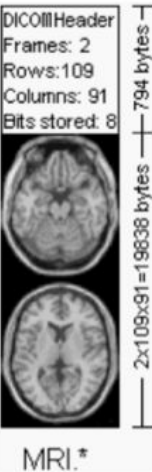


FIGURE 4.6 – Interactions dans la plateforme COOVADIS

L'extraction des informations des diagnostics n'est pas une tâche difficile car la norme DICOM déclare dans des entités spécifiques les données qui sont classées par segment (par exemple un segment ST peut traduire un infarctus du myocarde en phase aiguë). De plus, la norme est construite autour d'un langage standardisé facilement interopérable par les systèmes d'information (Figure 4.7). L'avantage d'utiliser un standard est qu'il existe de nombreux outils qui facilitent la gestion et la manipulation des messages sans qu'il

soit nécessaire de développer un système de traitement automatique des langages.

Pour la tâche de classification de connaissances, nous avons utilisé *dcm4che* qui permet d'étudier le corpus textuel et les concordances des expressions. Cette application créée en *Java* permet d'extraire toutes les expressions nécessaires pour notre traitement ontologique (les phases de diagnostic et de traitement, les concepts, les coordonnées du patient, les coordonnées du médecin et de l'institution responsable, ainsi que la manufacture et l'image médicale). Ces séquences sont reconnues d'après les patrons lexique-syntaxique que nous avons construits pour alimenter les ontologies.



Nom du champ	Balise	Contenu
▼ DICOMObject		
MetaElementGroupLength	0002,0000	206
FileMetaInformationVersion	0002,0001	
MediaStorageSOPClassUID	0002,0002	1.2.840.10008.5.1.4.1.1.4
MediaStorageSOPInstanceUID	0002,0003	1.2.840.113619.2.135.3596.3078086.5010.1206687344.107
TransferSyntaxUID	0002,0010	1.2.840.10008.1.2.1
ImplementationClassUID	0002,0012	1.2.276.0.7238010.5.0.3.5.4
ImplementationVersionName	0002,0013	OSIRIX
SourceApplicationEntityTitle	0002,0016	SCANPLUS2
Unknown	0008,0000	418
SpecificCharacterSet	0008,0005	ISO_IR 100
▼ ImageType	0008,0008	
value		ORIGINAL\PRIMARY\OTHER
value		ORIGINAL
value		PRIMARY
value		OTHER
SOPClassUID	0008,0016	1.2.840.10008.5.1.4.1.1.4
SOPInstanceUID	0008,0018	1.2.840.113619.2.135.3596.3078086.5010.1206687344.107
StudyDate	0008,0020	20080328
SeriesDate	0008,0021	20080328
AcquisitionDate	0008,0022	20080328
ContentDate	0008,0023	20080328
StudyTime	0008,0030	131253.000000
SeriesTime	0008,0031	131253.000000
AcquisitionTime	0008,0032	131253.000000
ContentTime	0008,0033	131253.000000
AccessionNumber	0008,0050	
Modality	0008,0060	MR
Manufacturer	0008,0070	GE MEDICAL SYSTEMS
InstitutionName	0008,0080	IRM DE NORD FRANCHE COMTE
ReferringPhysiciansName	0008,0090	
StationName	0008,1010	GEMSGEMS
StudyDescription	0008,1030	CAI
SeriesDescription	0008,103e	REP. 3 PLANS
ManufacturersModelName	0008,1090	SIGNA EXCITE
Unknown	0009,0000	156
Unknown	0009,0010	GEMS_IDEN_01
Unknown	0009,1002	GEMS
Unknown	0009,1004	SIGNA
Unknown	0009,1027	1206709973
Unknown	0009,1030	X117934001
Unknown	0009,1031	9999
Unknown	0009,10e3	1.2.840.113619.1.135.5.3596.3078086
Unknown	0009,10e7	2397840500
Unknown	0009,10e9	1206709973
Unknown	0010,0000	90
PatientsName	0010,0010	jean
PatientID	0010,0020	1918
PatientsBirthDate	0010,0030	19260106
PatientsSex	0010,0040	F
PatientsAge	0010,1010	082Y
PatientsWeight	0010,1030	79
AdditionalPatientHistory	0010,21b0	

FIGURE 4.7 – Méta-données d'une image DICOM

4.2.1/ COMMUNICATION ENTRE LES ONTOLOGIES

Compte tenu de la diversité de nos 3 ontologies (sur le domaine représenté, le vocabulaire, la représentation des connaissances et des axiomes), il est nécessaire d'établir un lien conceptuel entre elles afin de pouvoir les utiliser dans la même application (cf. Figure 4.3). Une alternative possible aurait été de développer une ontologie globale (une base de connaissances générale pour le domaine modélisé). Cependant, notre objectif était

de développer des ontologies réutilisables :

- l'ontologie de pathologies (ontologie du système vasculaire dans le premier prototype), qui est une ontologie qui pourra être réutilisée par exemple pour un système expert d'aide au diagnostic,
- l'ontologie dédiée au diagnostic (qui est un diagnostic d'interprétation sur le standard DICOM), et l'ontologie élaborée pour le travail collaboratif entre médecins (ontologie de traçabilité). Ces deux ontologies pourront être réutilisées avec une nouvelle ontologie de pathologie d'un autre domaine de santé.

Nos trois ontologies ont été élaborées pour permettre le couplage de celles-ci, c'est-à-dire qu'il existe des concepts de relation entre les ontologies pour former un réseau d'ontologies. De tels réseaux peuvent prendre la forme de réseaux pair-à-pair sémantique pour associer les connaissances décrites. Pour simplifier l'interconnexion des ontologies, nous avons fait le choix d'utiliser *l'union d'ontologies*¹ qui est fait pour l'ensemble d'un index I conceptuel (processus effectué avec l'outil *Jena*), qui est composé par la formule suivante :

$$\left[(O_i)_{i \in I}, (A_{ij})_{(i,j) \in I^2} \right] \cup \left[(O'_i)_{i \in I}, (A'_{ij})_{(i,j) \in I^2} \right] = \left[(O_i \cup O'_i)_{i \in I}, (A_{ij} \cup A'_{ij})_{(i,j) \in I^2} \right]^*$$

Où :

O correspond à l'ontologie,

A correspond au concept atomique,

I correspond à l'index conceptuel,

(i, j) correspond aux coordonnées hiérarchiques de l'ontologie

L'union ontologique a été faite entre l'ontologie du système vasculaire et l'ontologie du diagnostic, qui partagent des concepts index référant à la pathologie du patient (concept général de la hiérarchie ontologique : *vascular disorder*, *heart disorder*, *brain disorder*, *peripheral disorder*, *renal disorder*). De même, l'union ontologique entre l'ontologie du diagnostic et l'ontologie de traçabilité qui partagent des concepts index fait référence au patient et au dossier médical (concept général de la hiérarchie ontologique : *Patient et Study ID*) (cf. Figure 4.8).

L'étape de traitement consiste à appliquer les patrons lexicaux qui peuvent être reconnus par *dcm4che* dans la structure DICOM. Ainsi, ils permettent d'analyser des segments textuels porteurs de sens et de les étiqueter sur les graphes syntaxiques d'ontologies manipulées par *Jena*. Les séquences extraites correspondent aux patrons de l'ontologie alimentée. Par exemple, une phrase complète obtenue par analyse qui représente les instances peut être : *Stenosis in the left coronary artery, injured by cigarette smoking*. Nous pouvons éventuellement attribuer des sorties aux transitions du graphe comme : < Stenosis,.Affection >, < left coronary artery,.ObjetAnatomique >, < coronary,.ObjetAnatomiqueCoeur >, < injured,.Qualificatif > et < cigarette smoking,.OrigineDeAffection >.

Les ressources lexicales sont regroupées par une forme associative, soit en forme dérivée, soit en forme synonyme. Ces formes sont récupérées en accord avec le dictionnaire ontologique par le terme préféré du concept (preferred term), lequel est un concept

1. Union des ontologies, <http://www-kasm.nii.ac.jp/~koide/SWCLOS2/Manual/13OWLEntailments.htm>

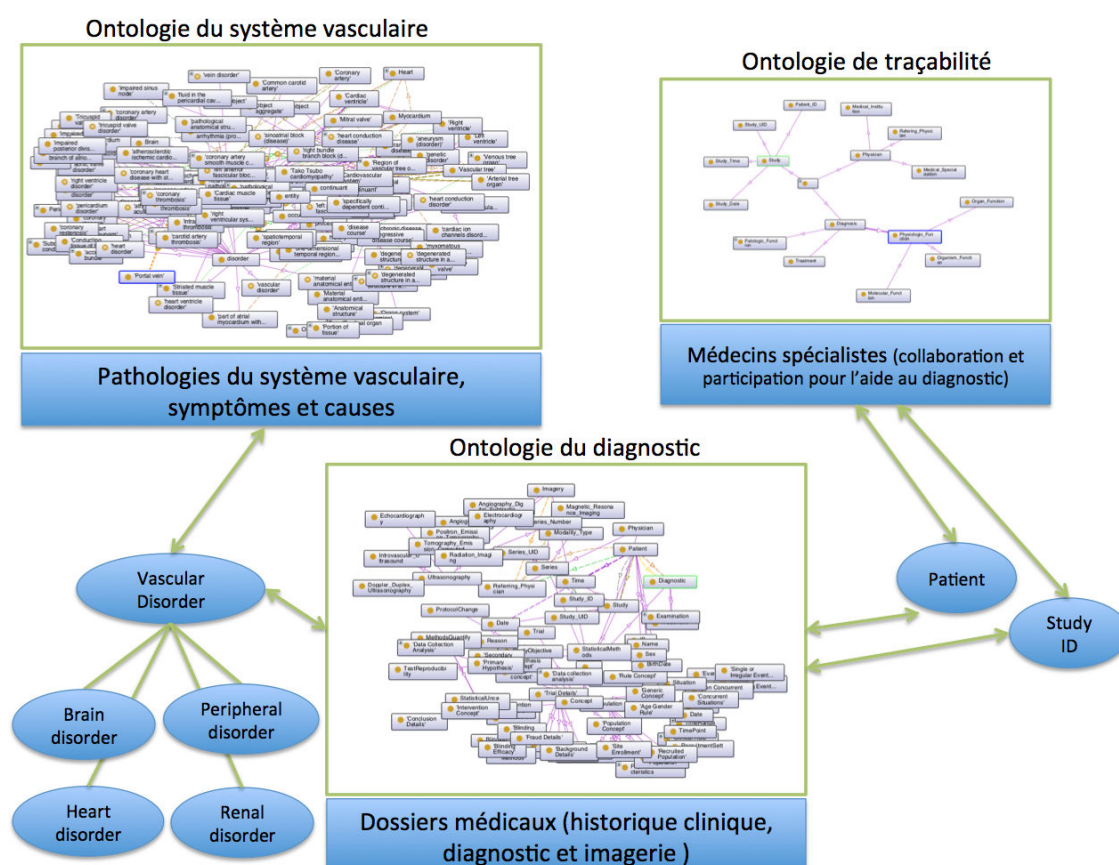


FIGURE 4.8 – Union des ontologies

représenté par son ID et Terme provenant du MeSH (Figure 4.9).

Les concepts définis peuvent être composés de deux ou plusieurs concepts primitifs reliés entre eux par une ou plusieurs relations. La plupart du temps, une expression est modélisée par un concept unique grâce à la structuration de la norme DICOM qui permet d'indexer l'information bien structurée et groupée comme observé dans la Figure 4.7. Il est possible que des concepts présentent de multiples relations sémantiques (par exemple *the infarction was caused by the buildup of the plaque in the coronary artery*, ou *the brain infarction by fatty material on the walls in right carotid artery*, ou *heart infarction as a result of slow blood flow and oxygen to the heart...*). La relation explicite de ces liens doit être faite pour le codage médical défini dans les ontologies.

Les concepts définis dans le diagnostic respectent les contraintes *Pathologie-Observations médicales - Collaboration pour l'union ontologique* qui permettent de retrouver les instances présentes et partagées entre les ontologies.

Dans nos choix d'organisation, nos ontologies peuvent effectivement intégrer des concepts distribués dans le domaine de la pathologie, *diagnostic-traitement* et travail collaboratif grâce à la *fusion ontologique* (opération d'union des ontologies) qui était nécessaire pour la modélisation correcte des connaissances pour la plateforme collaborative. Nous avons conçu un modèle structurel composé de 6 sections principales pour la représentation des connaissances (cf. Figure 4.10) : ❶ Pathologies (description des divers dysfonctionnements vasculaires du patient), ❷ Phénotypes (description

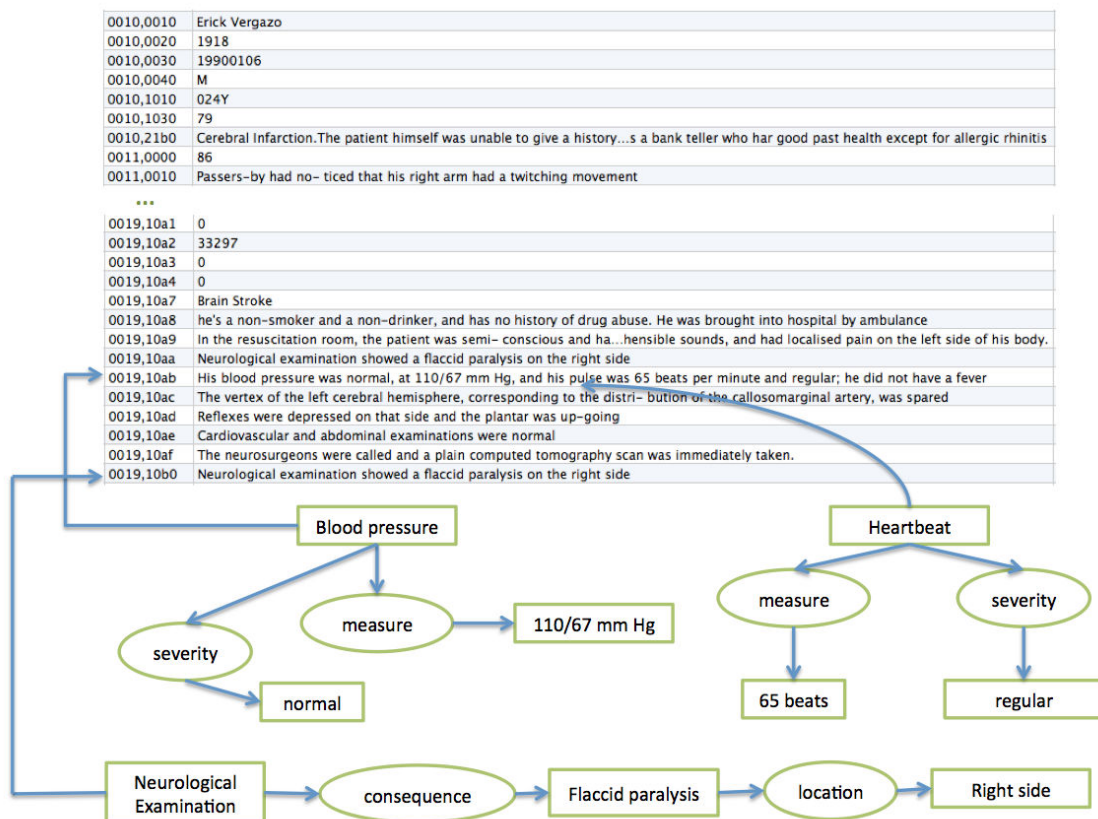


FIGURE 4.9 – Exemple d'association des concepts

des symptômes et des causes du patient), **③** Traitements (description générale de tous les traitements pour les maladies dont souffre le patient), **④** Procédures (description des procédures chirurgicales ou procédures d'exploration physiques faites au patient), **⑤** Pratique médicale (énumération des médecins et des institutions médicales qui ont traité le patient), et **⑥** Collaboration (énumération des médecins qui ont participé dans le travail collaboratif pour une prise de décision, ainsi que l'affichage de leur participation individuelle).

4.3/ FONCTIONNALITÉS DE COOVADIS

La plateforme COOVADIS est un système médical pour structurer et représenter la connaissance de diagnostic du système vasculaire. Cette plateforme reflète les relations médicales d'un patient, des médecins traitants et de leurs diagnostics sur l'évolution des pathologies du patient. La Figure 4.11 montre la page principale du portail web. Dans cette page, les médecins visualisent quatre sections : **①** le menu du portail, sur lequel le médecin peut changer de page pour visualiser toute l'information liée au diagnostic ou pour faire une nouvelle requête, **②** les images de présentation qui illustrent les institutions médicales (Centro Union et UPAC) et les instituts et Universités (Université de Franche-Comté, Université du Mexique UMSNH et institut Femto-st) qui ont contribué à

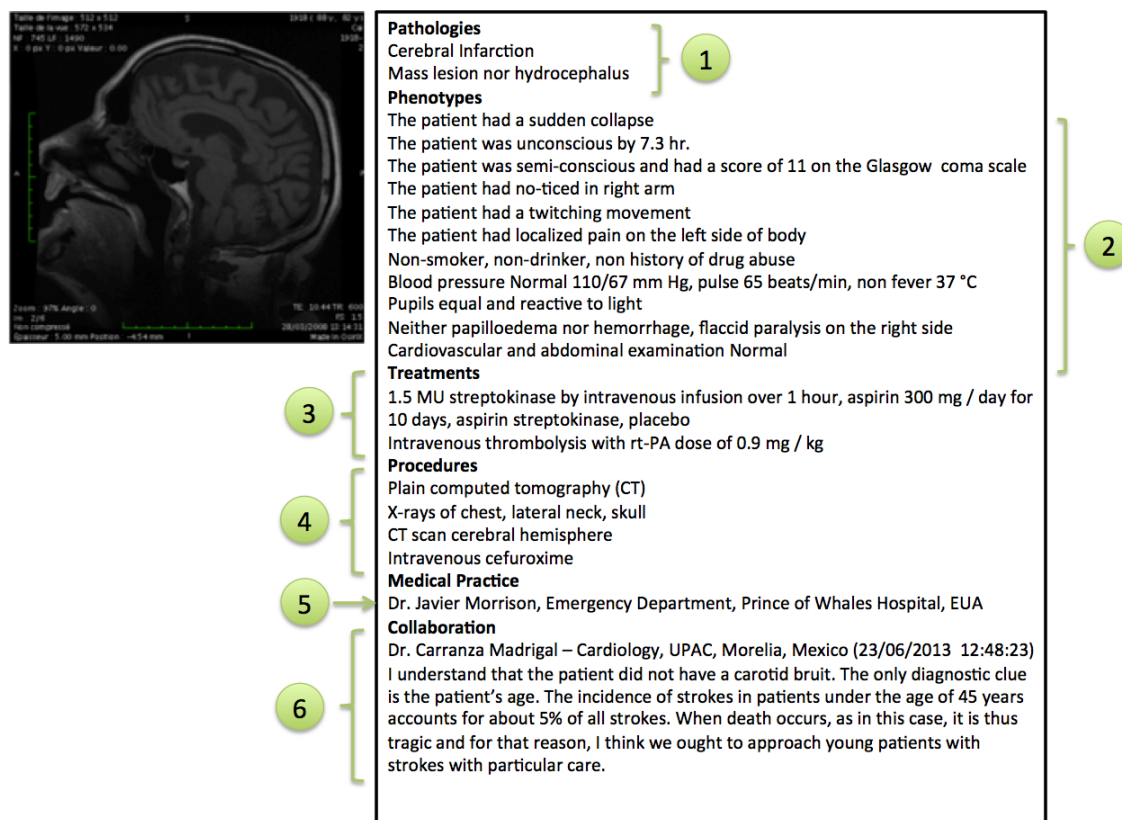


FIGURE 4.10 – Représentation des connaissances intégrées par les ontologies

ce projet, **3** la description des fonctionnalités de COOVADIS sous forme d'illustrations et de textes, et **4** la consultation d'informations accessibles par requête à l'aide de trois boîtes de texte, pour lesquelles il faut saisir les données correspondant au patient (numéro d'identification de patient et numéro d'identification du médecin).

La figure 4.12 montre la page d'affichage des connaissances liées au diagnostic obtenues grâce aux ontologies. Cette page est composée des réponses ontologiques et elle est structurée également en 4 sections : **1** le nom du patient ou les références de la requête pour la consultation médicale, **2** l'imagerie DICOM qui correspond au répertoire des dossiers médicaux représentés par les images d'exploration faites du patient, **3** l'information générale qui énumère l'information liée à chaque image DICOM, telle que les données générales du patient, du médecin traitant, du médecin radiologue ou technicien, du type d'image exploratoire et du fabricant (type de machine d'exploration clinique et marque de la machine), et **4** la connaissance sémantique des diagnostics obtenus à partir des deux ontologies groupées en 5 principales sections de représentation des connaissances (pathologies, phénotypes, traitements, procédures et collaboration).

Enfin, la Figure 4.13 montre la page de collaboration au diagnostic qui permet aux médecins de partager leurs observations, leurs expériences ou leurs expertises pour l'aide à la réalisation du diagnostic. Cette page est divisée en deux parties : **1** l'ajout d'informations personnelles, sur laquelle les médecins indiquent leurs coordonnées (nom complet, carte professionnelle de santé CPS3, spécialité, institution de santé), et **2** la boîte textuelle pour la participation au diagnostic.

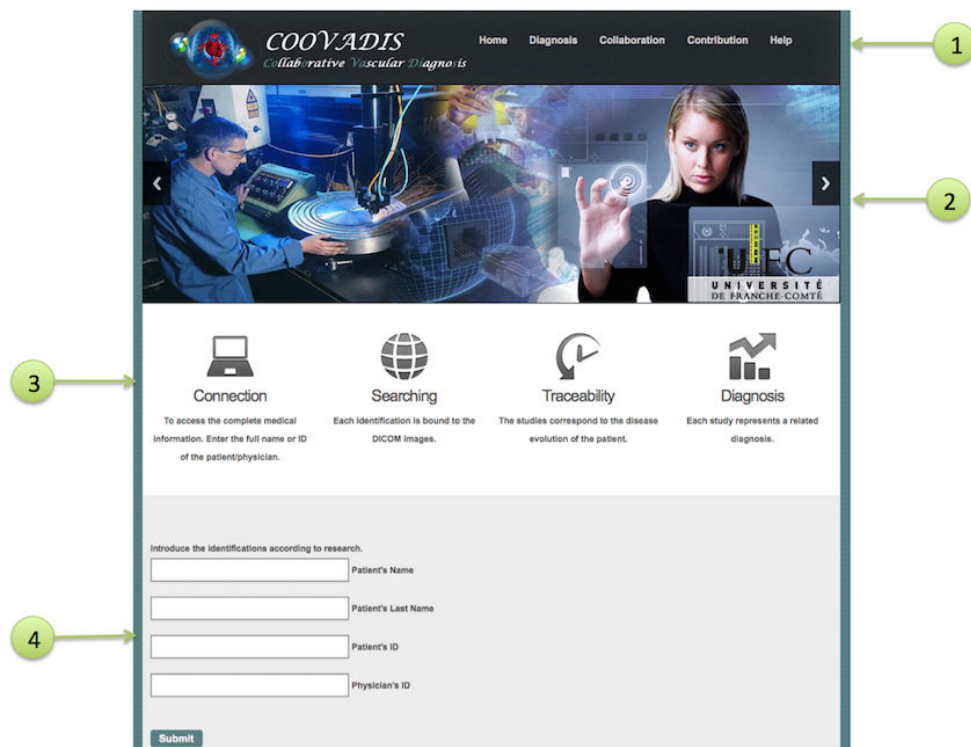


FIGURE 4.11 – La page principale du portail web

4.4/ EXEMPLE D'UN CAS CLINIQUE

Nous présentons dans cette section deux dossiers médicaux d'un patient réel. L'enrichissement des dossiers médicaux (dans les images au format DICOM) a été effectué par le médecin radiologue Dr. Fernando Sánchez du Centro Unión, au Mexique. Il détaille les diagnostics résultant d'imageries faites sur patient. Le Dr. Jaime Carranza Madrigal, médecin cardiologue de l'hôpital Civil de Morelia, a complété ces informations par l'entretien avec patient, ainsi qu'à l'aide des diagnostics résultant de l'examen physique et du traitement.

Cas clinique élaboré le 3 avril 2008 :

Mr J., âgé de 64 ans (Figure 4.14 se plaint au réveil à 7h d'une chute du membre supérieur gauche. Il remarque des fourmillements dans le bras gauche. À la marche, son genou gauche a tendance à se dérober et il sent moins bien le sol de ce côté. Il prend comme traitement pour de l'hypertension artérielle du COTAREG 160/12 et du LODOZ 10 et pour un diabète du DAONIL 5 et du GLUCOPHAGE 1000. Cet homme consomme régulièrement de l'alcool : de 3 à 5 verres par semaine. Il pèse 110kg pour 1.73 m. Il fume 10 cigarettes par jour.

L'examen physique :

Il existe une chute du bras G au maintien des attitudes, une chute modérée du membre inférieur G. La force musculaire de la main G est diminuée. La préhension aveugle du pouce gauche par la main droite est approximative, le sens de position du gros orteil G est mal perçu. Il existe une quadranopsie inférieure G. La tension est à 210 /110, le pouls

COOVADIS
Collaborative Vascular Diagnosis

Home Diagnosis Collaboration Contribution Help

PATIENT DIAGNOSIS

1 Patient's Name: aydee sanchez
Last study: Mon Aug 04 16:11:37 CEST 2014

2 Medical Studies

Tue Aug 28 00:00:00 CEST 2007
Wed Aug 29 00:00:00 CEST 2007
Thu Aug 30 00:00:00 CEST 2007
Tue Aug 28 00:00:00 CEST 2007

Diagnostic
November 11, 2012

Taille de l'image: 512 x 512
Taille de la vue: 572 x 620
M: 745 U: 1490
X: 0 px Y: 0 px Valeur: 23.00

1918 (88 V, 82 V)
Ca
1918

Zoom: 112% Angle: 0
Im: 216
Non compressé
Dossier: 5.00 mm Position: 7.52 mm

TE: 10.44 TR: 600
PS: 1.5
28/03/2008 11:14:11
Made in Ours

3 Patient Information

Patient Name: NINO MARTINEZ JOAQUIN
Patient ID: 280807-90617-AM
Birth Date: Tue Aug 16 00:00:00 CEST 1938
Birth Time: 000000
Sex: M

Physician Information

Referring Physician Name: DR. CARRANZA
Institution Name: CENTRO UNION
Institution Department Name: ULTRASONIDO
Performing Physician Name: DR. SANCHEZ
Operator Name: FSC

Study Information

Image Type: ORIGINAL
Study Date: Tue Aug 28 00:00:00 CEST 2007
Series Date: 20070828
Content Date: 20070828
Series Time: 091058
Content Time: 091342
Modality: US
Manufacturer: GE Medical Systems
Station Name: LOGIC9
Manufacturer Model Name: LOGIQ9

4 Pathologies

Cerebral infarction
Mass lesion nor hydrocephalus

Phenotypes

The patient had a sudden collapse
The patient was unconscious by 7.3 hr
The patient was semi-conscious and had a score of 11 on the Glasgow coma scale
The patient had no-ticed in right arm
The patient had a twitching movement
The patient had localised pain on the left side of body
Non-smoker, non-drinker, non history of drug abuse
Blood pressure Normal 110/67 mm Hg, pulse 65 beats/min, not fever 37°C
Pupils equal and reactive to light
neither papilloedema nor hemorrhage, flaccid paralysis on the right side
Cardiovascular and abdominal examinations Normal

Treatments

Procedures

plain computed tomography (CT)
X-rays of chest, lateral neck, skull
CT scan cerebral hemisphere
Intravenous cefuroxime

Medical Practice

Dr. Javier Morrison, Emergency Department, Prince of Wales Hospital, EUA

Collaboration

Coovadis: Thesis Project - Femto, Université de Franche-Comté

FIGURE 4.12 – La fenêtre d'affichage des connaissances liées aux diagnostics

est rapide et irrégulier.

L'examen d'imagerie :

La topographie de fibres pyramidales et de fibres proprioceptives indique une Quadranop-sie inférieure gauche qui résulte d'une atteinte des radiations optiques pariétales droites.

L'examen Doppler des vaisseaux du cou montre une sténose carotidienne.

Diagnostic :

Le patient présente une neuropathie diabétique. La TDM (Tomodensitométrie) réalisée sans injection montre un effacement des sillons corticaux dans la région gauche et une lacune dans le striatum droit. Le patient présente un AVC ischémique sylvien droit.

FIGURE 4.13 – La fenêtre de collaboration au diagnostic

L'échantillon d'ECG (électrocardiogramme) montre : TP(Taux de prothrombine) de 9.12, TCA(Temps de céphaline activée) de 33.67, Troponine (IDM indolore chez diabétique) Bilan lipidique = cholestérol total 120 mg/dl, HDL-cholestérol 32 mg/dl protéinurie 4.07, microalbuminurie bilan HTA (Hypertension artérielle) de 150 mg/dl Il est recommandé d'administrer de l'aspirine (dose de charge : 160 - 325 mg) dans les 48 heures après l'AIC (Accident Ischémique Constitué) (Grade I, niveau A)

Sept mois plus tard, le 24 novembre 2008, le patient a présenté un deuxième accident vasculaire cérébral. Le cas est le suivant :

Mr J., âge de 64 ans, fumeur, porteur d'un diabète de type 2 révélé il y a 7 ans. Le patient présente brutalement un déficit moteur du côté gauche responsable d'une chute. Ses antécédents personnels résident en une hypertension artérielle traitée par du COTAREG 160/12, du LODOZ 10 et du BISOPROLOL 10mg, et des chiffres de PA élevés à plusieurs occasions, fluctuant entre 150/94 et 162/104 mmHg. Il précise qu'au cours des 10 dernières années des chiffres élevés de PA avaient été retrouvés de façon inconstante. Il existe un antécédent d'AVC ischémique avec sur-poids. Le poids actuel est de 115 kg pour 1.73 m. Le périmètre abdominal est de 147 cm.

L'examen physique :

Le patient est conscient, il présente un déficit moteur du côté gauche à 2/5 avec troubles sensitifs de même topographie et difficultés d'expression orale avec paraphasies. Il

présente une HTA de grade 1 à 2 chez un diabétique de type II, fumeur. Il s'agit d'un sujet à risque cardiovasculaire élevé en référence à l'âge, sexe, présence d'un diabète de type II et un antécédent d'AVC ischémique. La TA est de 180/100 mm de mercure. Le pouls est irrégulier et l'ECG montre un tracé de FA. Sur le plan anamnestique, il s'agit d'un sujet hypertendu traité. L'auscultation cardiaque est normale de même que la palpation des pouls périphériques. L'examen vasculaire n'a pas retrouvé d'anomalie et en particulier pas de souffle sur les trajets artériels. Le pouls est à 88 battements/minute, et la température est de 37.4 degrés.

L'examen d'imagerie :

L'IRM (L'imagerie par résonance magnétique nucléaire) montre des déficits sensitifs ou moteurs unilatéraux, une aphasie et une cécité monoculaire transitoire. Le patient présente un syndrome optico-pyramidal, un déficit brachio-facial et un trouble du langage avec un déficit du membre supérieur dominant. Ils sont le fait d'un déficit carotidien.

Diagnostic :


La glycémie suggère un contrôle glycémique insuffisant, mais une Hb glyquée est indispensable. Il existe une dyslipidémie associée au diabète II avec hypertriglycémie, un HDL bas et un LDL à 1.45 g/l (formule de Friedwald) ; il s'agit des anomalies lipidiques liées à l'insulinorésistance à situer dans le contexte de diabète 2. On retient la présence d'une microalbuminurie significative à 60 mg/24h. Le traitement initial comprend un anti-hypertenseur assurant une néphroprotection par Losartan 50-100mg 1cp et Atorvastatine 10mg.

A la suite des examens, la première étape consiste à récupérer l'ensemble des informations présentes au sein de l'image DICOM, c'est-à-dire, les informations personnelles du patient (ID, nom, âge, sexe...) ainsi que le diagnostic (ID de l'image DICOM, type d'imagerie, date de création...). Cette étape permet ainsi d'adapter le contenu présent au sein des images au format de l'ontologie afin d'en extraire uniquement les données pertinentes pour les besoins des professionnels de la santé. La seconde étape consiste en l'interprétation des interventions cliniques chez le patient (les types d'études, les observations des résultats et les traitements). Cette étape permet d'intégrer l'ensemble des dossiers d'un patient au sein d'un même système facilitant ainsi la visualisation des diagnostics et l'accès aux informations.

La connaissance liée au patient est traitée par un processus de filtrage et de restructuration basé sur quatre objectifs principaux :

- (1) fournir une description du patient (historique clinique),
- (2) déterminer les acteurs (médecins) directement impliqués dans la prise en charge du patient,
- (3) fournir une description intégrée des maladies et traitements du patient pour la compréhension de certains aspects de l'évolution du patient,
- et (4) produire une structure de connaissances intégrée dans la plateforme afin d'analyser l'état du patient.

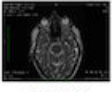
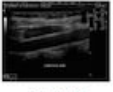
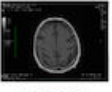
Nous savons qu'il est difficile dans un système de terminologie de retrouver le terme qui exprime fidèlement le sens associé au libellé d'un diagnostic. L'utilisation de nos ontologies va permettre de lever ces ambiguïtés. Chaque concept est désigné par un terme préféré qui correspond au terme employé par le médecin (le terme préféré est la forme canonique du terme). À partir de l'ontologie, nous avons pu extraire automatiquement


COOVADIS
Collaborative Vascular Diagnosis

[Home](#)
[Diagnosis](#)
[Collaboration](#)
[Contribution](#)
[Help](#)

PATIENT DIAGNOSIS

Patient's Name: FRANCISCO J. JUAREZ MARTINEZ
Last study: Mon Nov 24 00:00:00 CEST 2008

Medical Studies




Sat Apr 03 00:00:00 CEST 2008
Sat Apr 03 00:00:00 CEST 2008
Mon Nov 24 00:00:00 CEST 2008

Diagnostic
November 24, 2008

Taille de l'image: 512 x 512

Taille de la vue: 572 x 427

NF: 888 LF: 1777

X: 0 px Y: 0 px Valeur: 0.00


AS

1918 (24y, 18y)

Ca

1918-

3



Zoom: 83% Angle: 0

Im: 3/6

Non compressé

TE: 88.192 TR: 3200

FS: L5

28/03/2008 13:17:55

Study Information
Image Type: ORIGINAL
Study Date: Sat Apr 03 00:00:00 2008
Series Date: 20080403
Content Date: 20080403
Series Time: 124052
Content Time: 131524
Modality: US
Manufacturer: GE Medical Systems
Station Name: LOGIC9
Manufacturer Model Name: LOGIQ9

Padecimientos
Hipertensión Arterial Sistémica
Diabetes Mellitus tipo II
Neuropatía Diabética
Accidente Cerebral Vascular de tipo Ataque Isquémico Transitorio localizado
Arteria Cerebral Media
Fenotipos
Fumador, Bebedor, No Historia de drogas
Repentino colapso de brazo izquierdo Extremidad Izquierda
Disminución muscular
Cuadrantanopsia Inferior Izquierda
Control Glucémico Inadecuado
Dislipidemia Mellitus Insulino Dependiente, Hiperfrígliceridemia
Análisis Lipídico Colesterol Total 120mg/dl Lipoproteína de Alta Densidad 32mg/dl
Proteinuria 4.07
Hipertensión 150mg/dl, Presión Arterial 210/110
Tratamientos
Losartan Sustancia Activa 50mg
Atorvastatina Sustancia Activa 10mg
Aspirina Sustancia Activa 160mg
Procedimientos
Electrocardiograma: Tiempo de Protrimbina 9.12, Tiempo de Coagulación Activado 33.67
Tomodensitometría Sin inyección: Supresión de surcos corticales Localización
Región Izquierda Cerebro, Laguna localización Región derecho
Doppler Carotida: Estenosis
Imagen de Resonancia Magnética: Déficit Sensorial, Afasia
Práctica Médica
Dr. Jaime Carranza, Cardiology, UPAC, Morelia, Mexico
Colaboración
Dr. Carranza Madrigal – Cardiology, UPAC, Morelia, Mexico (12/05/2014 18:08:51)
El paciente debe de tener una dieta estricta y una actividad regulada, deberá de eliminar el cigarro, alcohol y cafeína en su dieta

Coovadis- Thesis Project - FEMTO, Université de Franche-Comté

FIGURE 4.14 – Saisie du dossier de Mr J.

les informations nécessaires pour la construction d'une sémantique à savoir les termes préférés associés aux concepts de l'ontologie et leurs synonymes.

Un concept défini peut être composé de deux ou plusieurs concepts primitifs reliés entre eux par une ou plusieurs relations. La plupart du temps, une expression est modélisée pour un concept défini. Cette combinaison sous-entend la présence de relations entre ces concepts, même si elles ne sont pas explicitées. Une définition explicite de ces liens doit être faite pour le codage DICOM. Ces liens vont être définis dans l'ontologie. La formation des concepts définis provient d'un mode de composition des concepts et des relations régi par des contraintes. Par exemple, l'expression *Hypertension artériel* est modélisée par un concept défini appartenant à l'axe *Pathologie*.

Les instances de patrons textuelles repérées dans le texte (diagnostics dans DICOM)(Figure 4.15) sont traitées de façon à pouvoir les afficher par une expression reconnue comme étant un diagnostic. Précisément, lors de l'application des patrons, chaque terme reconnu est précédé de sa catégorie. Nous recherchons alors la forme canonique, définie dans les ontologies, de chacun des termes reconnus et des relations existantes entre eux (Figure 4.16).

Comme nous l'avons précédemment indiqué, nos ontologies sont en espagnol pour permettre une évaluation par des cardiologues mexicains de Morelia, et en anglais afin d'être également utilisables en Europe. Les dossiers médicaux sont en espagnol en raison de l'origine des diagnostic (Hôpital Civil, Centro Union et UPAC, qui sont des centres médicaux mexicains).

Cas Clinique du 5 avril 2008

0019,10a7	El paciente presenta una neuropatía diabética
0019,10a8	El TDM sin inyección muestra supresión de surcos corticales en la región Izq y una laguna en el estriado Der
0019,10a9	
0019,10aa	ACV isquémico en ACM
0019,10ab	ECG muestra TP 9.12 TCA 33.67
0019,10ac	CT 120mg/dl HDL 32mg/dl proteinuria 4.07
0019,10ad	HAS 150mg/dl
...	
0019,10b7	Administración de aspirina 160 mg después de 48hr al AI

Cas Clinique du 23 novembre 2008

0019,10a7	IRM muestra un déficit sensorial de motores unilateral, una afasia y ceguera monocular transitoria
0019,10a8	El paciente presenta un síndrome de óptica-piramidal, un déficit braquio-facial, lenguaje con un déficit de la extremidad superior dominante
0019,10a9	El paciente presenta déficit de la carótida
0019,10aa	La glucemia indica un control glucémico inadecuado
0019,10ab	Dislipidemia asociada a DMID tipo 2 con hipertrigliceridemia
0019,10ac	HDL bajo, LDL 1.45g/l
0019,10ad	microalbuminuria 60 mg/24h
...	
0019,10b7	Administración de antihipertensivo LOSARTAN 50-100 1 capsula y ATORVASTATINA 10mg

FIGURE 4.15 – Méta-données des dossiers médicaux en image Dicom

Lexique Espagnol/Anglais (lexique simplifié permettant au lecteur de comprendre la Figure 4.16 : *Accidente Cerebral Vascular/Cerebral Vascular Accident, Afasia/Aphasia, Análisis/Analysis, Arteria Cerebral Media/Middle Cerebral Artery, Ataque Isquemico Transitorio/Transient Ischemic Attack, Colesterol Total/Total Cholesterol, Déficit Sensorial/Sensory Deficit, Derecho/Right, Estudio/Study, Hipertensión Arterial Sistémica/Systemic Arterial Hypertension, Imagen/Image, Imagen de Resonancia Magnética/Magnetic Resonance Imaging, índice/Rate, Izquierdo/Left, Medida/Measure, Laguna/Lacuna, Lípidico/Lipidic, Lipoproteína de Alta Densidad/High-Density Lipoprotein, Localización/Location, Observación/Observation, Padecimiento/Disease, Perfil/Profile, Proteinuria/Proteinuria, Región/Region, Sustancia Activa/Active Substance, Sin*

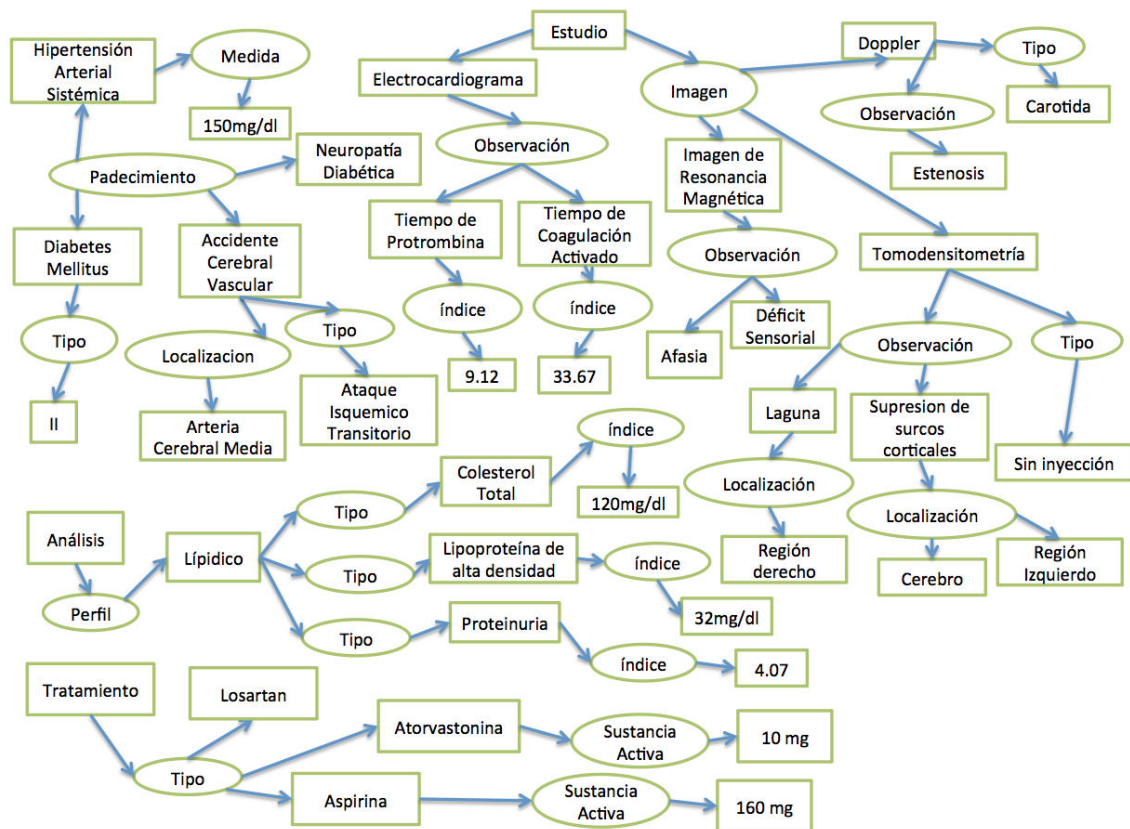


FIGURE 4.16 – Concepts et relations

Inyección/No injection, Tipo/Type, Tratamiento/Treatment.

Nous avons développé une plateforme qui permet de construire la connaissance sémantique sur des diagnostics à partir des ontologies. Nous nous servons des principaux axes conceptuels du domaine qui ont été préalablement identifiés dans les ontologies. Les actions médicales, les signes, les pathologies, les examens, les traitements... Le système compare les termes médicaux extraits : si dans le thésaurus modélisé nous avons la conjonction de termes, alors une correspondance est observée, récupérée et affichée dans l'interface utilisateur (cf. Chapitre 5 Section 5.4).

CONCLUSION



Nous avons présenté notre plateforme web pour le travail collaborative appelé COOVADIS (*COllabOrative VAscular DiagnoSis*) : cette plateforme utilise les trois ontologies définies dans le chapitre 3 (l'ontologie du système vasculaire, l'ontologie du diagnostic et l'ontologie de la traçabilité). Nous avons présenté une partie plus technique dans la section 4.2 sur l'implémentation que nous avons réalisée, enfin nous avons exposé le premier prototype de COOVADIS.

Précisons que nous avons dû réellement finaliser ce premier prototype, car c'est grâce à lui que les médecins ont pu tester ses fonctionnalités, son utilisabilité, et la souplesse et l'efficacité apportées par l'utilisation de nos ontologies.

Nous avons terminé ce chapitre par quelques explications basiques agrémentées de copies d'écran sur ce premier prototype qui a permis aux médecins d'effectuer les premiers tests. Ainsi sur un dossier existant le cas de Mr J. (qui nous servira de *fil rouge* dans le chapitre suivant) nous avons pu exposer les résultats fournis par notre plateforme.

Le chapitre suivant est consacré à la validation théorique et à la validation à l'aide d'un test en milieu clinique. Pour aborder cette partie il était nécessaire au préalable de proposer ce premier prototype.

VALIDATIONS EN MILIEU CLINIQUE ET PREMIERS TESTS EXPÉRIMENTAUX

INTRODUCTION

Ce chapitre s'intéresse à la validation et à l'évolution de nos ressources termino-ontologiques dans le domaine du diagnostic des pathologies vasculaires et du travail collaboratif. Elles permettent de mettre à disposition les connaissances présentes au sein des diagnostics au travers d'une plateforme collaborative pour faciliter le travail entre professionnels (*cf.* Chapitre 4). La combinaison des sources ontologiques et d'une plateforme sont nécessaires pour couvrir le spectre complet des décisions médicales et de leur justifications dans le contexte d'étude.

Nous avons effectué deux types de validation :

- la validation par outils informatiques qui vérifie la syntaxe des ontologies, elle permet d'éliminer quelques erreurs telles que la redondance ou l'omission, l'incohérence, l'erreur de partitionnement.
- la deuxième validation a été effectuée par un expert qui vérifie la cohérence de l'ensemble des connaissances et de la sémantique.

Dans ce chapitre après la description du cycle de vie des ontologies, nous verrons comment faciliter les interactions avec les experts pour cette validation (l'expert n'étant pas familier avec les langages informatiques) puis quels sont les types de validation à effectuer (pour l'expert et les outils). Nous décrirons ensuite les modifications apportées à nos ontologies et terminerons par les premiers résultats expérimentaux à l'aide de l'exemple *fil rouge* de Mr J. (*cf.* Section 4.3.1.) : ces résultats ont été validés (validation en milieu clinique) par deux professionnels de santé.

5.1/ CYCLE DE VIE DES ONTOLOGIES

Dans le cycle de vie d'une ontologie et de l'usage de l'ontologie, il existe différentes phases de conception et de développement(*cf.* chapitre 2 section 2.3 cycle de vie). Le processus de développement est composé de phases classiques (besoins ou spécification, conceptualisation, contrôle, formalisation, analyse, assurance qualité par la validation d'expert, mise en œuvre et maintenance), ces phases suivent un modèle itératif composé de boucles et de cycles entre ces phases qui permettent l'analyse et la ré-validation avant la mise en œuvre et l'usage. Le processus cyclique de vie dans une ontologie peut être vu comme un cycle composé pour 4 étapes : ❶ identifier des objec-

tifs et alternatives de conception ontologique qui correspondent à des activités de gestion de projet (planification) ❷ évaluer les alternatives et identifier les risques potentiels (contrôle et assurance qualité), ❸ développer et valider les tâches de conceptualisation (intégration, acquisition des connaissances, spécification, conception, formalisation et analyse), et ❹ valider les résultats. Ces étapes sont composées d'un ensemble de phases successives ou parallèles (Figure 5.1), où à chaque phase est attribué à un rôle bien défini.

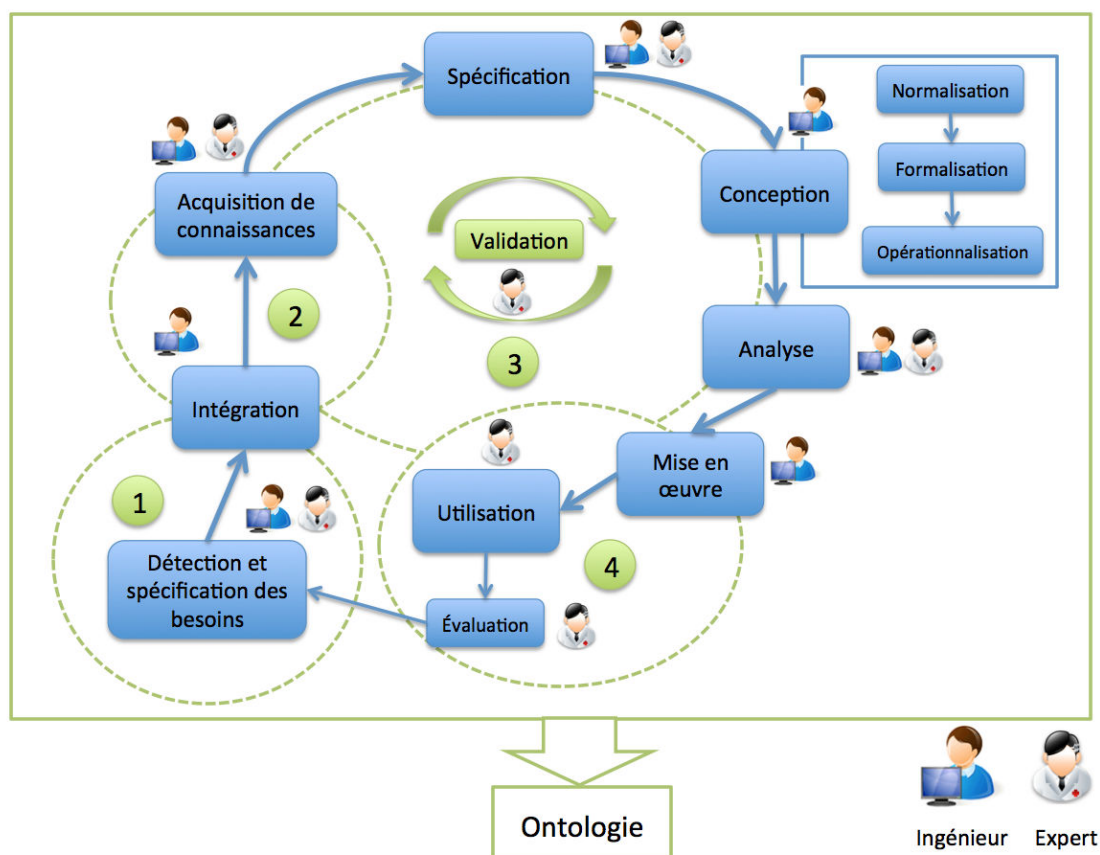


FIGURE 5.1 – Le processus cyclique de vie dans une ontologie

Détection et spécification des besoins. Les objectifs étaient de permettre le partage d'informations provenant des images DICOM pour l'amélioration du diagnostic entre professionnels de la santé. Ces professionnels peuvent connaître les références cliniques en fonction de plusieurs variables, telles que les maladies, les symptômes, les médicaments, les diagnostics... déjà faits en extrapolant l'évolution du patient pour améliorer la pratique clinique et la prise de décisions, grâce au travail collaboratif. Cela représente la phase d'analyse du problème à résoudre et le point d'entrée du modèle ontologique. Grâce à l'utilisation des ontologies nous obtenons l'identification et l'analyse d'un domaine global sur un même patient au sein du même système : COOVADIS, plateforme d'exploitation d'informations médicales et de collaboration entre médecins. Cette phase a été réalisée en collaboration avec des médecins afin de connaître quels sont leurs besoins et leurs attentes quant au système proposé. Les résultats obtenus de cette analyse sont les suivants :

- Une connaissance des traitements médicaux spécifiques qui ont été réalisés sur un patient ;
- Une étude comparative des différents traitements sur le patient avec les caractéristiques mesurables de la progression de la maladie ou le risque ;
- Des événements défavorables associés aux traitements médicaux spécifiques ;
- Un espace de collaboration pour travailler sur un plan de traitement plus optimal en fonction de l'historique obtenu du patient via l'extraction de données médicales.

Intégration. Dans cette phase, nous avons cherché des modèles ontologiques existants qui correspondent aux critères de la phase d'analyse du problème. Malheureusement, nous n'avons pas trouvé un modèle qui comprenne les associations nécessaires pour le diagnostic de maladies vasculaires et leur traçabilité. Pour ce dernier point, nous avons développé les ontologies qui résolvent ce problème (cf. chapitre 3).

Acquisition de connaissances. Cette phase implique principalement l'interaction par l'analyse de textes et de manuels, des outils taxonomiques et des entrevues formelles et informelles avec des experts. Elle favorise la conception du vocabulaire et des définitions, relations, conditions d'interaction . . . pour discerner toutes les connaissances nécessaires sur plusieurs ressources afin de déterminer un modèle ontologique approprié à plusieurs niveaux de conceptualisation. La phase d'acquisition comprend l'utilisation des glossaires, des groupes de tables pour englober : types de concepts, synonymes et acronymes, classes, associations, attributs et instances. Cette phase d'acquisition des connaissances peut être effectuée en parallèle avec les phases de spécification, de conception et d'analyse.

Spécification. Cette phase est une étape très importante dans le cycle de vie du développement d'une ontologie, phase similaire au développement d'un logiciel. Dans cette phase l'ingénieur ontologique interagit avec les utilisateurs finaux (les médecins) afin de définir la portée et les détails fonctionnels du modèle d'ontologie (phase d'analyse du problème), les associations (phase d'intégration) et la connaissance critique d'application (phase d'acquisition de connaissances). Dans cette phase, il est possible d'initier la validation et si nécessaire de spécifier des changements. En général, la spécification demande un effort de collaboration entre les acteurs impliqués dans le développement et la validation.

Conception. Cette phase est le cœur du cycle de vie de développement où l'ingénieur ontologique identifie les concepts impliqués dans la connaissance pour la classification (analyse de textes), l'agrégation ou le changement sur des niveaux de conceptualisation (en accord avec les résultats de la phase de spécification pour le *design pattern* ontologiques) ce qui correspond à la normalisation. La formalisation des connaissances se situe à un certain niveau d'abstraction, lequel correspond à la représentation conceptuel sur la logique de description (TBox et ABox). Finalement, pour l'opérationnalisation nous avons utilisé la représentation sémantique OWL.

Analyse. Cette phase est l'étape d'itération qui doit être exécutée avant la mise en œuvre de l'ontologie parce que cette phase valide les modifications et les résultats atten-

dus par les experts. Durant la conception et l'analyse, les connaissances sémantiques nécessaires sont vérifiées afin de s'assurer de leur exactitude. Par exemple, l'expert teste les modèles ontologiques avec leurs questions et il vérifie aussi si le problème général d'analyse a été abordé.

Mise en œuvre et utilisation. Dans le domaine des ontologies, l'étape finale est la mise en œuvre et l'utilisation à l'aide d'une plateforme ou d'un outil d'exploitation. L'utilisation peut être aussi un processus cyclique, s'il y a une identification des termes, du vocabulaire ou d'ontologie qui ne répond pas aux besoins du domaine.

La difficulté des ontologies est souvent liée au type d'environnement dynamique conceptualisé par de multiples acteurs impliqués et par le domaine distribué où l'ontologie va travailler. Elles ne peuvent être pensées comme une conceptualisation finie d'un domaine de connaissances délimité et stable. Les motifs de changements dans une ontologie sont multiples :

- le domaine de définition peut évoluer ce qui implique des modifications dans l'ontologie pour rendre compte de l'évolution ;
- la modification de l'application au sein de laquelle est utilisée l'ontologie implique la nécessité d'adapter l'ontologie ;
- la conceptualisation même de l'ontologie peut être affinée à travers le processus cyclique de validation. Ainsi pour répondre aux objectifs organisationnels et d'usage, les ontologies doivent s'adapter aux besoins.

5.2/ VALIDATION DES ONTOLOGIES

5.2.1/ INTERACTIONS AVEC LES EXPERTS

Le développement et l'adaptation des ontologies médicales sont des tâches complexes qui exigent un effort humain considérable et une étroite collaboration avec des experts médicaux [Asm13]. Même si les outils et/ou les techniques de développement des ontologies sont assez complets pour soutenir ce travail, la complexité du langage formel dans le domaine médical et les techniques d'ingénierie informatique nécessitent une collaboration constante entre ces deux profils (l'ingénieur et le médecin). En effet, les experts du domaine médical ne sont généralement pas familiarisés avec les langages informatiques et d'ingénierie de connaissances. Cette collaboration requiert beaucoup de temps pour éviter des erreurs de compréhension entre les deux acteurs.

La validation des aspects logiques et structurels de l'ontologie (incohérence, omission ou redondance des concepts) peut être réalisée automatiquement à l'aide d'outils comme Protégé ou Jena [Hua08], mais ces outils fournissent seulement des solutions partielles qui demandent des interventions manuelles de l'ingénieur ontologique. D'autre part, la validation de la conceptualisation est plus complexe à faire et il est nécessaire de demander l'aide des experts médicaux.

Pour valider la conceptualisation des ontologies par des experts médicaux ne possédant pas de connaissances en informatique, nous avons utilisé une approche de questions-réponses interactives et de diagrammes pour simplifier notre logique de description ontologique (Figure 5.2). Cette tâche comprend deux étapes principales :

- la création de questions en langage naturel sur la hiérarchisation des termes et leur logique de connexion,
- l'interprétation des informations fournies par les experts pour déduire si la partie de l'ontologie qui a été évaluée est valide, invalide ou a besoin de modifications.

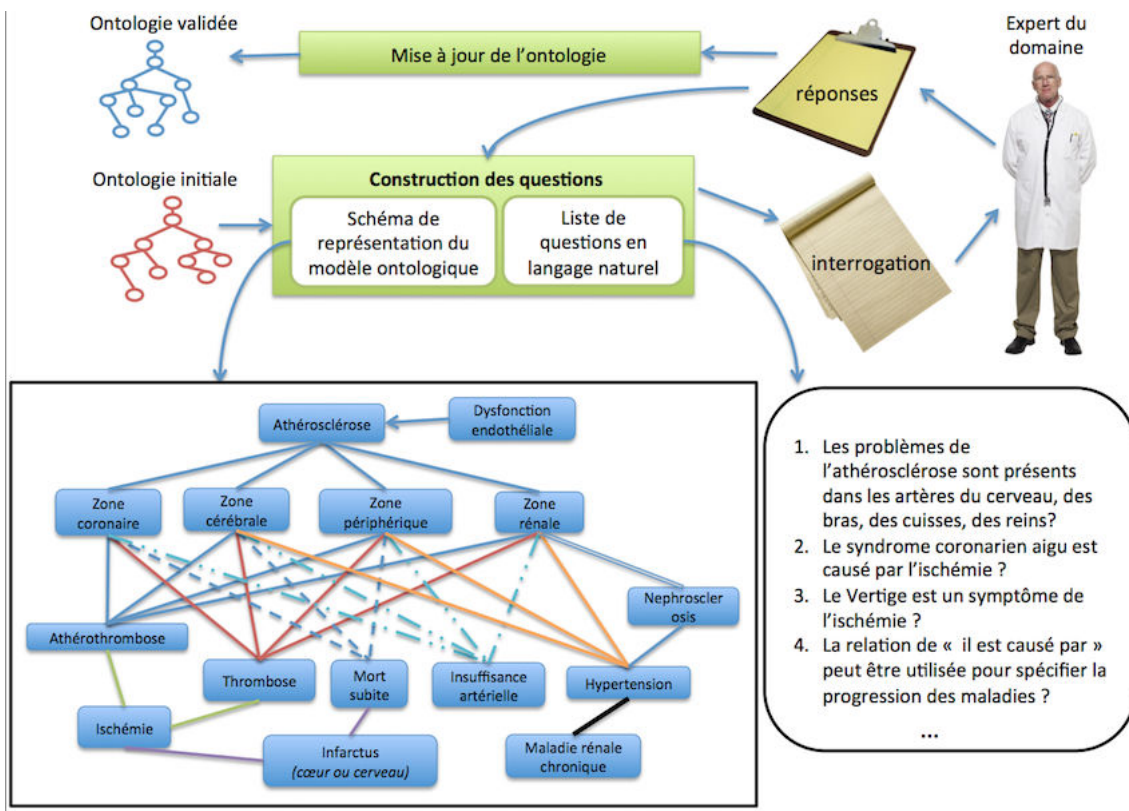


FIGURE 5.2 – Validation de la conceptualisation ontologique pour une approche de questions-réponses interactives

Les critères de validation de l'ontologie définissent un ensemble de primitives de représentation avec pour objectif de modéliser le domaine de connaissances. Les primitives de représentation sont typiquement les classes, les attributs et les relations. Les définitions des primitives de représentation incluent des informations sur leur signification et les contraintes de leur application logique et cohérente. Plusieurs critères sont utilisés pour la validation des ontologies, certains d'entre eux traitent de l'exactitude formelle de l'ontologie comme par exemple :

- Les erreurs de duplication : certains éléments de l'ontologie peuvent être déduits à partir d'autres concepts.
- Les erreurs de disjonction : une classe est définie comme une conjonction de classes distinctes.
- Les erreurs de consistance et cohérence : les définitions actuelles conduisent à des conclusions contradictoires.

Un autre type de critère de validation est la complétude. Gómez-Pérez dans [Gom07] fait remarquer qu'il est impossible de prouver la complétude d'une ontologie, mais il est possible de prouver le caractère incomplet d'un élément de l'ontologie. Dans ce contexte, une autre question importante est la validation au niveau des connaissances du domaine

de l'ontologie en raison de l'énorme quantité d'informations qui doit être modélisée. Par exemple, dans le domaine médical, le nombre de connaissances est doublé tous les cinq ans voire même tous les deux ans [Asm12].

La validation de nos ontologies a été réalisée en utilisant des schémas de connaissance ontologique et de description de leurs individus. Plus précisément, nous nous concentrons sur le niveau de conceptualisation du domaine en répondant à quatre questions principales :

1. Quels éléments doivent être validés ?
2. Comment classer les éléments à valider ? Quelles validations sont indépendantes ? Quels éléments sont dépendants les uns des autres ?
3. Comment valider ces éléments ?
4. Comment faire les mises à jour nécessaires après chaque étape de validation ?

Plusieurs critères peuvent être utilisés pour évaluer la qualité de la conceptualisation du domaine. Nous nous concentrons en particulier sur cinq critères :

1. l'application du vocabulaire,
2. la correspondance de la taxonomie (la généralisation et la hiérarchisation),
3. l'adéquation des relations non-taxonomique, c'est-à-dire, l'ajustement des relations sémantiques,
4. la cohérence et l'extensibilité, (L'ontologie doit être cohérente afin d'effectuer des déductions qui soient correctes, en accord avec les définitions disponibles. Elle doit être construite de façon à ce que tout ajout de concept n'affecte pas sa cohérence.)
5. la forme minimale de l'ontologie. (Elle doit avoir les hypothèses minimales sur le besoin à modéliser et ne doit pas contenir de connaissances supplémentaires qui ne sont pas utilisées dans la pratique actuelle.)

5.2.2/ TYPES DE VALIDATION

Pour analyser les besoins concrets de nos ontologies (ontologie du système vasculaire et ontologie de diagnostic), nous avons mené des entretiens avec des médecins spécialisés en cardiologie et radiologie pour valider nos ontologies médicales. La première étape a consisté en la construction d'une liste de questions en langage naturel à partir de l'ontologie à valider. Ces questions sont soumises à des experts du domaine qui offrent des réponses (réponses de validation, réponses de modifications ou d'ajout d'informations) en accord avec leur expérience et leur expertise. L'étape suivante consistait en l'interprétation des commentaires des experts pour valider ou modifier l'ontologie. Le processus proposé en Figure 5.2 basé sur des modifications-validations a été utilisé pour : (1) valider les ontologies construites basées sur des textes médicaux et sur des outils d'extraction terminologique (UMLS, ICD, MeSH, ...) ; et (2) valider la connaissance sémantique que les ontologies renvoient de la plateforme COOVADIS.

Le but de la première étape est de construire des connaissances formalisées dans les ontologies afin de valider un nombre maximum d'affirmations avec un nombre minimum de questions. Puisque nous avons utilisé des outils existants pour vérifier le formalisme ontologique (Protégé et Jena), ces outils ont validé les types de déclarations d'ontologie

en accord avec l'évaluation de redondance ou d'incohérence conceptuelle. Les questions sont construites sur des modèles qui représentent les éléments de l'ontologie. Elles sont du type suivant :

- Une classe A est une sous-classe de B, des questions accompagnées de schémas qui montrent la hiérarchisation de concepts,
- Les relations entre chaque classe de l'ontologie sont correctes,
- La propriété P est composée des propriétés de type Q,
- La classe B est du domaine de la propriété P,
- Un individu I appartient à la classe A,
- La propriété P relie les individus I et J.

Nous partons de l'hypothèse que tous les éléments d'une ontologie médicale doivent être validés. Il s'agit de concepts de types de maladies, de relations entre les maladies et les symptômes et les causes, d'instances de concepts qui permettent de connaître l'existence d'une pathologie, de relations entre les instances de concepts (par exemple une échographie artérielle peut être commandée pour la détection d'une dysfonction endothéliale) ou de relations entre les instances de concepts avec des littéraux (par exemple une phrase « donner par voie orale du charcoai 50g » indique la dose de la substance qui doit être administrée « 50g »). Ces éléments de l'ontologie indiquent les principaux mots-clés des questions qui sont liées aux concepts, aux instances et aux relations de l'ontologie.

L'évaluation de l'expert donne des réflexions menées autour des questions pour l'évolution de l'ontologie et de leur maintenance. La validation peut avoir plusieurs objectifs : effectuer un bilan d'ensemble de la représentation de la connaissance, guider la prochaine étape de construction et rendre compte de résultats. Certaines fois, les réponses des deux experts n'étaient pas simples et de nombreux échanges ont été nécessaires pour évaluer l'exactitude de la compréhension de leurs retours, mais également pour valider les modifications réalisées. La Figure 5.3 montre le cycle de validation.

Une ontologie, à l'instar d'une connaissance, doit être validée à plusieurs niveaux des ressources : (1) validation des corpus textuels utilisés pour le développement de l'ontologie ; (2) validation du contenu de l'ontologie ; (3) validation de la qualité taxonomique de l'ontologie ; (4) validation de l'ontologie dans son contexte d'usage et de sa réutilisation.

Validation du corpus textuel. Dans le développement d'une ontologie à partir de textes, de corpus, de ressources existantes telles que le méta-thésaurus UMLS et le thésaurus MeSH, le rôle des experts est d'identifier si les textes disponibles sont nécessaires ou pertinents en accord avec notre objectif. Dans notre cas, cela signifie : vérifier que les techniques ou concepts sélectionnés pour la conceptualisation correspondent bien à ceux qu'ils utilisent aujourd'hui.

Validation du contenu de l'ontologie. L'évaluation du contenu est basée sur la signification des termes définis en donnant des définitions objectives. L'ontologie doit être cohérente pour permettre des inférences conformes aux définitions. Plus spécifiquement, la validation du contenu ontologique est basée sur le modèle de structure et d'adaptation terminologique d'une connaissance en fonction des besoins. Plus particulièrement, le modèle doit s'attacher à représenter de manière explicite les connaissances sur des langages standardisés qui permettent la maintenance et l'extensibilité.

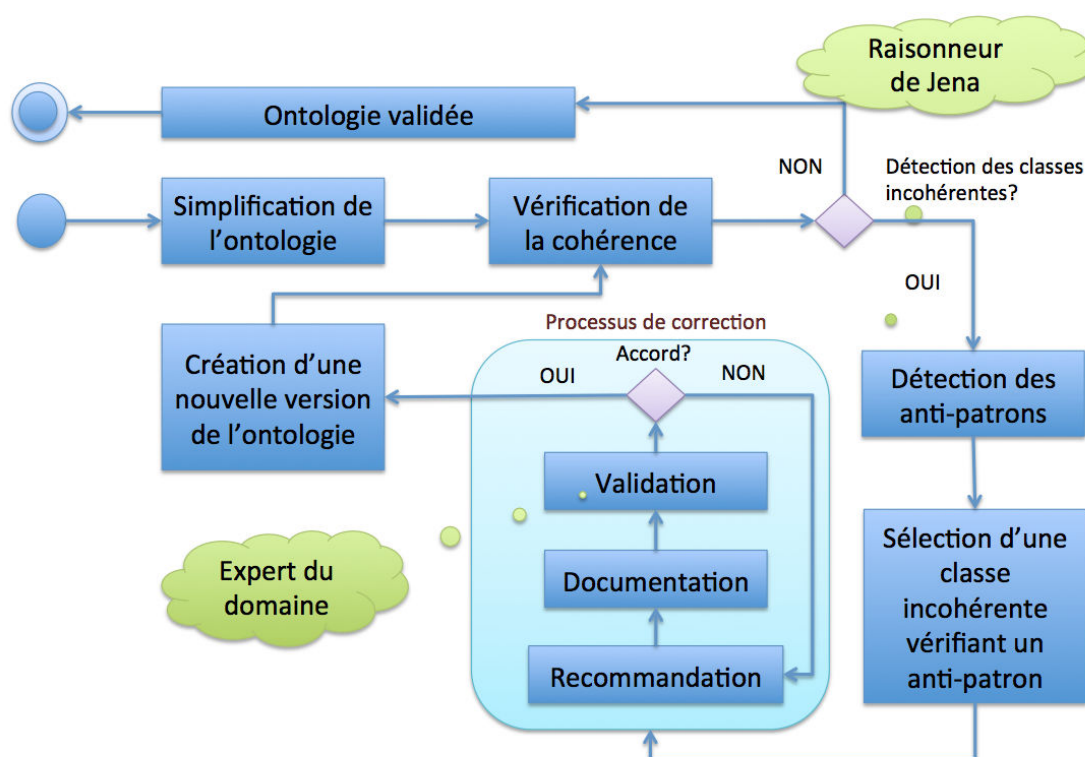


FIGURE 5.3 – Processus de validation

Validation de la qualité taxonomique de l'ontologie. Les ontologies sont organisées sous la forme d'une hiérarchie, cette organisation taxonomique s'impose pour structurer le modèle de connaissances. Le type de validation taxonomique est possible grâce à l'utilisation d'éditeurs d'ontologie (comme Protégé par exemple) et de raisonneurs (comme Jena par exemple) et par l'expert du domaine, qui vérifie l'exactitude de la structure. Les outils informatiques assurent que les définitions en langage naturel et en langage semi-formel respectent les spécifications de construction de l'ontologie, et l'expert assure une structuration correcte en accord avec son expertise. Les erreurs structurelles les plus communes sont les suivantes :

- les erreurs de circularité se produisent quand une classe est définie comme une spécialisation ou une généralisation d'elle-même : erreur détectable par les outils informatiques (Jena ou Protégé) ;
- les erreurs de partition qui peuvent apparaître quand la définition de la partition entre un ensemble de classes est omis (par exemple l'ontologie définit une partition d'une classe dans un ensemble de sous-classes qui ne sont pas disjointes et devraient l'être), erreur détectable par les outils informatiques ;
- les erreurs de redondance se produisent quand on redéfinit des expressions qui ont déjà été explicitement définies ou qui peuvent être déduites à partir d'autres définitions. Par exemple, un « infarctus du cœur » est un « infarctus du myocarde » mais c'est également une « crise cardiaque ». Dans cet exemple, l'erreur de redondance n'est pas découverte par la validation automatique de Protégé ou de Jena mais par l'expert du domaine. Ce type d'erreur ne permet pas de garantir une cohérence dans la hiérarchie de l'ontologie et peut introduire des erreurs quant aux résultats fournis par le système.

- les erreurs sémantiques sont produites habituellement quand l'ingénieur ontologique classe un concept comme sous-classe d'une classe à laquelle il n'appartient pas réellement, ces type d'erreurs sont détectables par l'expert.
- les erreurs d'incomplétude se produisent quand des concepts sont classés sans les expliquer complètement et quand l'expert du domaine ne trouve pas des concepts ou des relations nécessaires pour une structuration complète.

Validation de l'ontologie dans son contexte d'usage et sa réutilisation. Cette validation est l'étape la plus critique de l'ontologie. Dans cette étape, on valide la qualité et l'intérêt de l'ontologie. Cela correspond à vérifier que les résultats attendus par l'expert et les performances pour exécuter cette tâche sont correctes. L'expert médical qui est également l'utilisateur doit distinguer les critères d'efficacité suivants : (1) la pertinence du vocabulaire décrivant les concepts, (2) la pertinence de la hiérarchie des résultats, (3) la pertinence des relations sémantiques du contenu, (4) la détection des erreurs d'omission ou d'ajout d'informations attendues. Si l'ontologie réussit la validation d'usage par l'expert, il est possible de considérer que l'ontologie est réutilisable et peut être mise à disposition de la communauté.

5.3/ MODIFICATIONS DES ONTOLOGIES

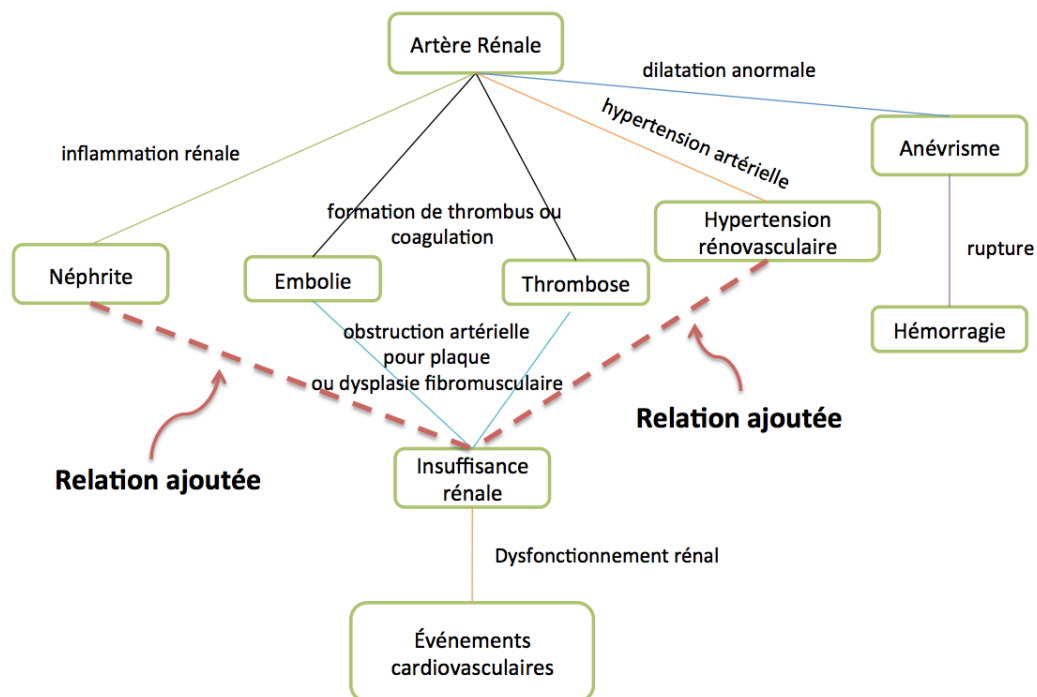


FIGURE 5.4 – Changement basique

La gestion des changements pour enrichir ou modifier l'ontologie est une tâche cruciale. L'ingénieur d'ontologie doit contrôler tous les effets des changements et évaluer l'impact du changement sur l'ontologie. Lors de la validation de l'expert du domaine sur nos on-

tologies (ontologie du système vasculaire et ontologie de diagnostic) celui-ci nous a proposé des changements sur des opérations basiques et sur des opérations complexes. Les opérations basiques incluent des changements simples et atomiques appliqués à une seule entité de l'ontologie (granularité fine). Les opérations complexes correspondent à des mécanismes regroupant des opérations basiques pour former des unités plus riches et de composition.

Dans un changement basique, les changements appliqués correspondent à des classes, des propriétés et des instances, qui sont un type de changement selon la taxonomie et les entités concernées. Pour ce type de changement, les contraintes à satisfaire doivent être appliquées en restrictions, cardinalité, domaine de propriété, échelle, disjonction ou jonction, symétrie, type . . . (Figure 5.4).

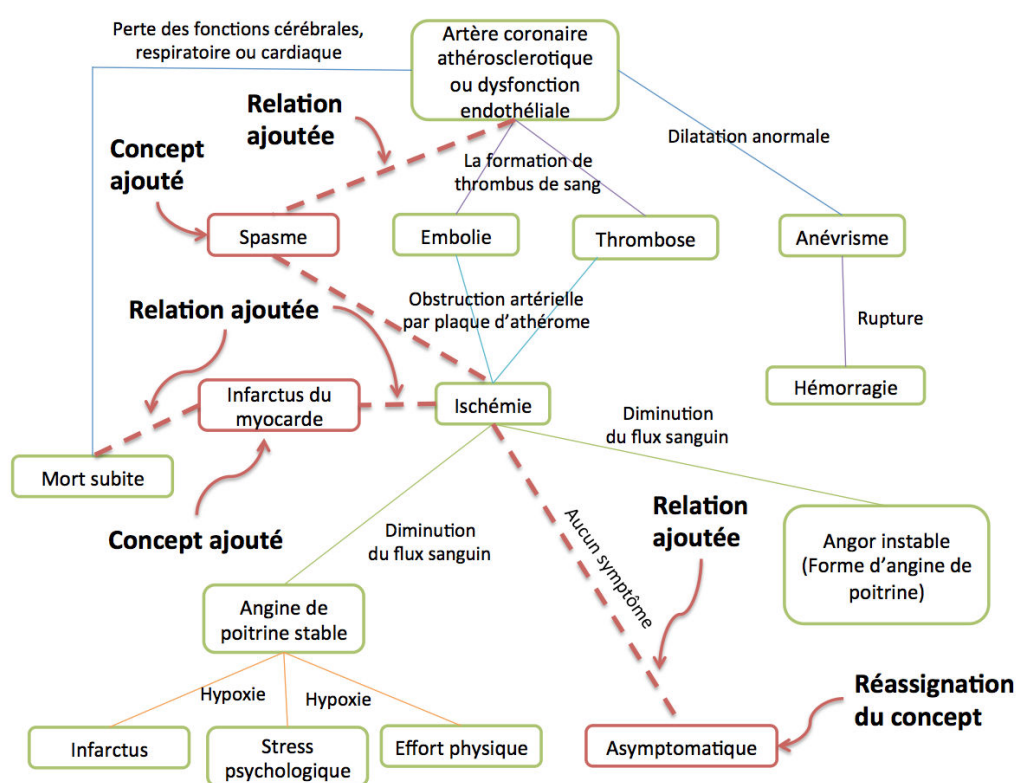


FIGURE 5.5 – Changement complexe

Dans un changement complexe, les changements appliqués correspondent à l'interprétation des concepts ou des relations, de telle sorte que de nouvelles règles ou connaissances doivent être modifiées ou recomposées. Par exemple, dans le domaine médical, des maladies peuvent avoir des conditions différentes ou des éléments spécifiques qui ne sont pas présents dans tous les cas cliniques. Les changements peuvent avoir plusieurs conséquences sur la connaissance de l'ontologie. Par exemple, la Figure 5.5 montre les types de changements réalisés sur une restructuration conceptuelle.

Les changements sont appliqués sur des versions de test de l'ontologie pour analyser leurs effets. La raison est qu'il est possible d'introduire des inconsistances de données en appliquant un changement (plus particulièrement lors de la suppression de composants dans l'ontologie). Il est important de maintenir l'ontologie dans un état consistant

en tenant compte du principe de la continuité ontologique, c'est-à-dire, tendre à minimiser les alternatives de suppression d'axiomes en proposant des alternatives de division de concepts, de fusion de concepts ou de redistribution d'instances. Dans notre changement, le type de suppression a été le suivant : le concept « Anterior Choroidal Artery Infarction » n'est pas considéré comme une condition ou implication du concept « Ischémie », les concepts « Infarctus cérébral » et « Mort subite » sont assignés au concept « Insuffisance artérielle » (Figure 5.6).

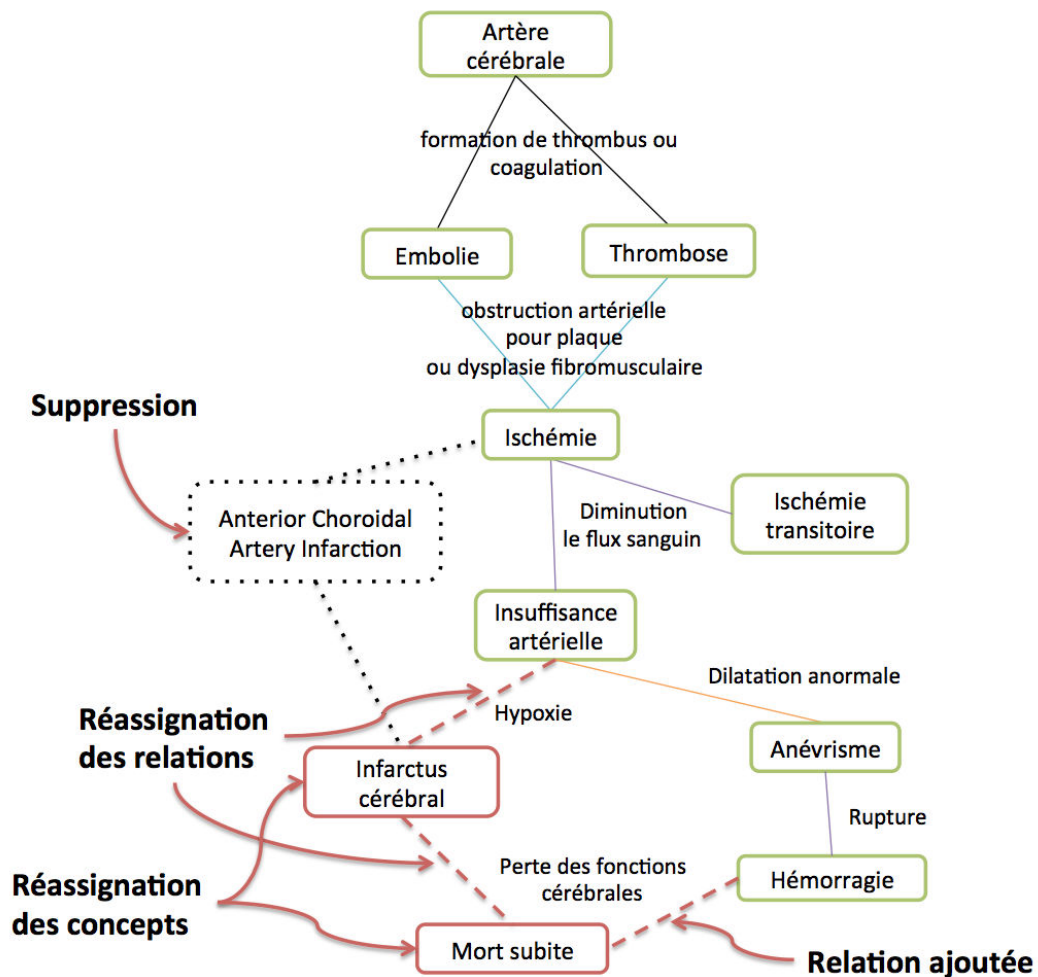


FIGURE 5.6 – Changement par suppression

Finalement l'application des changements correspond à la validation finale de l'ensemble des modifications apportées aux ontologies tout en s'assurant de la traçabilité de ces modifications. L'évaluation finale de l'expert a été faite en comparant les anciennes valeurs de l'ontologie initiale (sauvegardées dans les fichiers ontologiques d'évolution).

5.4/ PREMIERS TESTS DE COOVADIS

Dans le cadre de notre plateforme, il a été nécessaire de faire évaluer COOVADIS par un expert. Cette étape permet de vérifier que les résultats attendus sont corrects. Les critères de validation de l'expert sur la plateforme sont la pertinence des connaissances du do-

The screenshot displays the COOVADIS web application interface. The top navigation bar includes links for Home, Diagnosis, Collaboration, Contribution, and Help. The main content area is divided into several sections:

- Patient's Name:** Francisco J. HERRERA Martínez. Last study: Mon Nov 24 00:00:00 CEST 2008.
- Medical Studies:** Three small images showing brain scans with dates: Sat Apr 03 00:00:00 CEST 2008, Sat Apr 03 00:00:00 CEST 2008, and Mon Nov 24 00:00:00 CEST 2008.
- Diagnostic:** A large brain scan image with technical details: Taille de l'image: 512 x 512, Taille de la vue: 582 x 533, IM: 1000 LF: 2000, K: 0 px Y: 0 px Valeur: 31.00. Technical parameters include: Zoom: 89% Angle: 0, TE: 15.198 TR: 560, Im: 6/6, PS: 15, 28/03/2008 11:33:25.
- Diagnosis Contribution:** A section for inserting comments, including fields for Name, Specialty, and Institution, and a text area for comments.

The bottom of the page shows the footer: Coovadis - Thesis Project - Fentis, Université de Franche-Comté.

FIGURE 5.7 – Cas de Mr J. : Résultats de Diagnostic sur COOVADIS

maine (les concepts et relations sémantiques du contenu) et le formalisme qui est utilisé pour représenter les connaissances (le vocabulaire utilisé pour décrire les concepts).

Divers tests ont été appliqués pour évaluer la couverture des concepts ou des termes en cours d'utilisation. Nous avons analysé les erreurs et les imprécisions de notre plateforme et nous avons fait évoluer les ontologies pour améliorer nos résultats. L'analyse de ces résultats nous a aidé à enrichir nos ressources (ontologies) à base de synonymes et de termes non identifiés et qui ont été présents dans des diagnostics particuliers.

Dans notre plateforme, la représentation des connaissances des diagnostics (résultats de la plateforme) a été systématiquement soumise à validation récurrente. L'ambiguïté

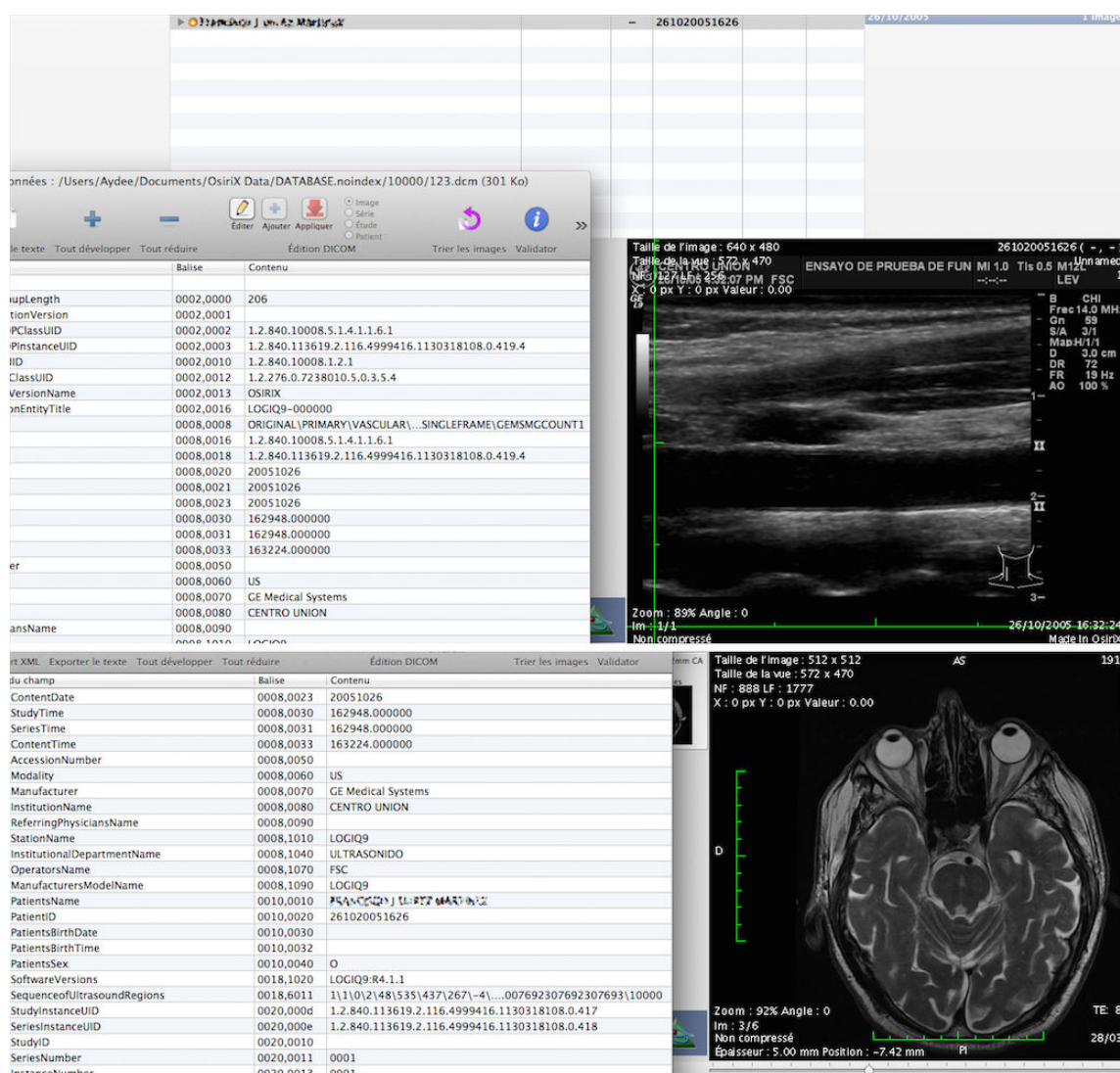


FIGURE 5.8 – Cas de Mr J. : Résultats avec enrichissement de métadonnées de diagnostique

de la langue (anglais et espagnol) nécessite que les ontologies possèdent un formalisme irréductible et intelligible par tous les utilisateurs. Dans ce contexte, les ontologies écrites dans les deux langues ont évolué en parallèle avec le thésaurus de spécialité. Cette évolution dépend des nouvelles dénominations des termes proposées par l'expert du domaine. Nous avons rencontré des situations qui ont nécessité des évolutions. Par exemple, les cas où des abréviations ou des acronymes sont utilisés par les médecins pour décrire des maladies, des causes, des parties du corps humain ou des organes. C'est le cas de *AF*-Antecedentes Familiares/Antécédents Familiaux, ou encore de *CII*-Cuadrante Inferior Izquierdo/Quadrant Inférieur Gauche et de *CII*-Carotida Interna Izquierda/Carotide Interne Gauche (cf. Chapitre 4 Figure 4.15). Dans ces exemples, les termes avaient une correspondance directe avec le concept de l'ontologie et les résultats des cas cliniques (cf. Chapitre 4 Section 4.4) et peuvent être visibles dans les Figures 5.7, 5.8, dans laquelle il est possible de voir une structuration de la connaissance validée par les experts : la description des divers dysfonctionnements vasculaires, des causes, des symptômes du patient, la description des procédures d'exploration faites au patient,

et la description des pratiques médicales pour la collaboration entre médecins.

La validation de la plateforme et des ontologies a été assurée, au Mexique, par les Docteurs Sonia Maria López Correa (Directrice de la clinique UPAC-*Unidad de Prevención y Atención Cardiometabolica*), Jaime Carranza Madrigal (de l'*Université Michoacana de San Nicolás de Hidalgo* et cardiologue à l'Hôpital Civil), et le Docteur Luis Fernando Sanchez Contreras (radiologue au *Centro Union de Diagnostico*) qui confirment leur participation et la validation de nos travaux dans les lettres : Figures 5.9, 5.10, 5.11, et 5.12

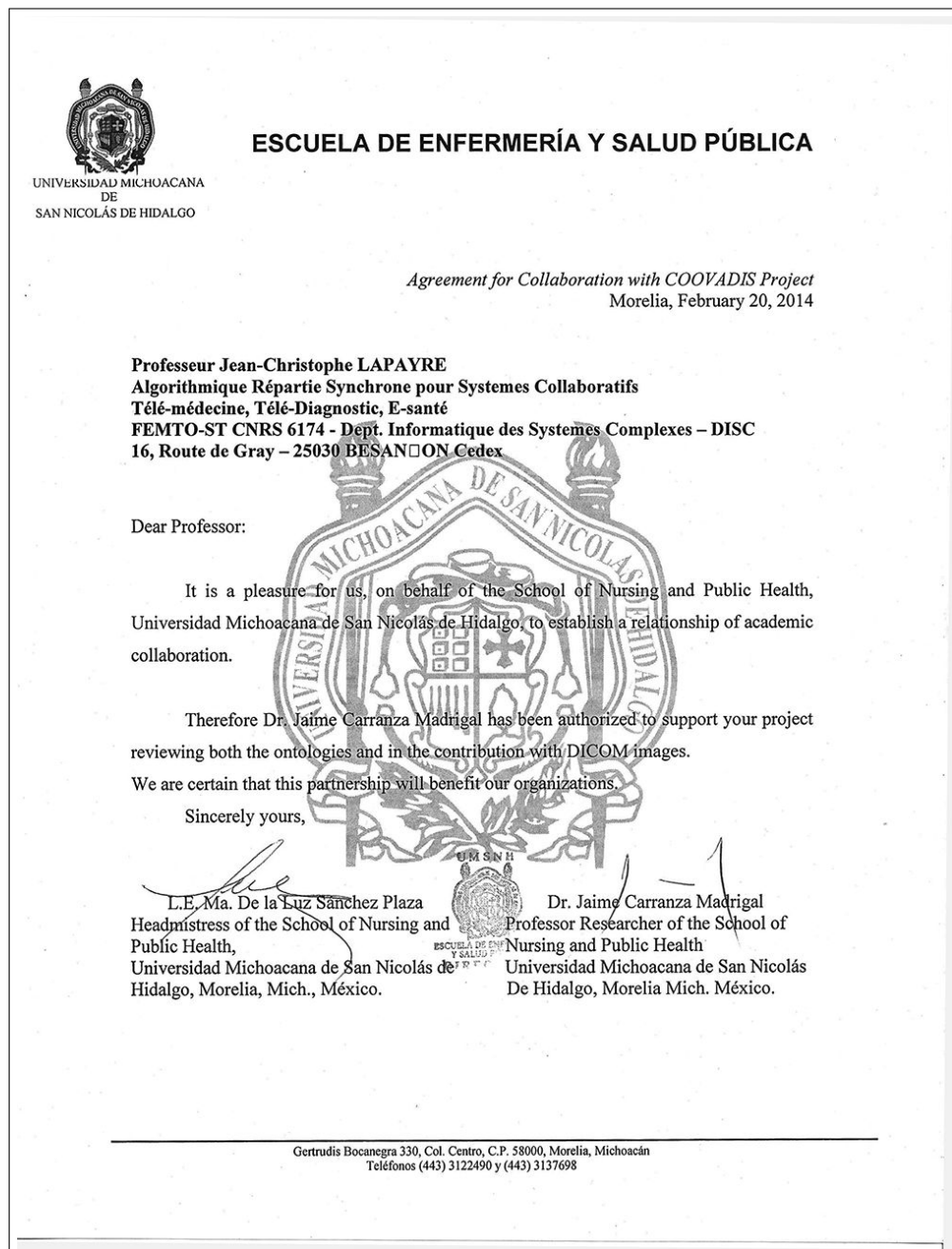


FIGURE 5.9 – Lettre de collaboration avec le Dr. Jaime Carranza Madrigal

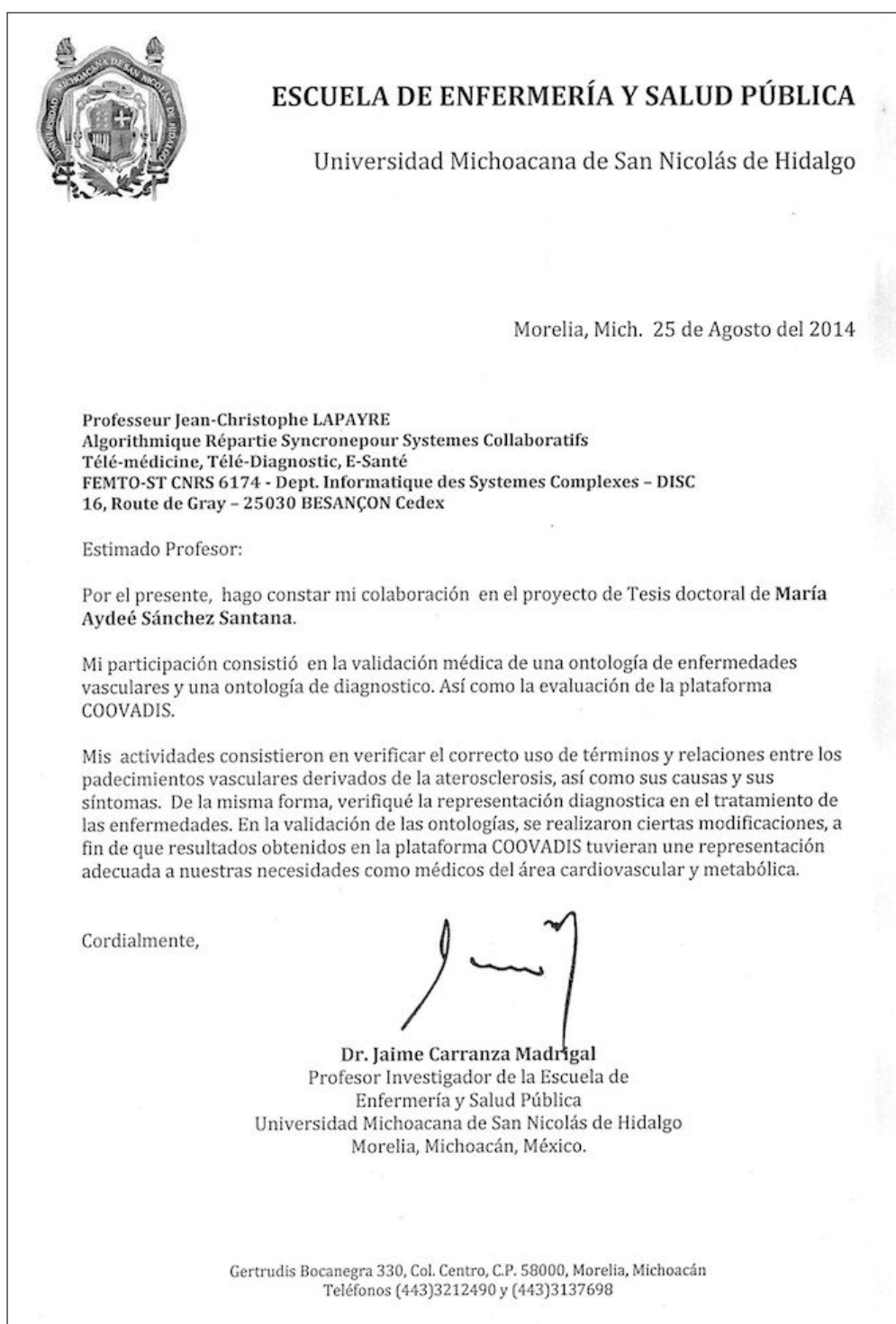


FIGURE 5.10 – Attestation du Dr. Jaime Carranza Madrigal pour la validation de COOVADIS



UPAC
UNIDAD DE PREVENCIÓN Y
ATENCIÓN CARDIOMETABÓLICA

Morelia, Michoacán, a 21 de Agosto del 2014

Professeur Jean-Christophe LAPAYRE
Algorithmique Répartie Synchron pour Systemes Collaboratifs
Télé-médecine, Télé-Diagnostic, E-Santé
FEMTO-ST CNRS 6174 - Dept. Informatique des Systemes Complexes – DISC
16, Route de Gray – 25030 BESANÇON Cedex

Estimado Profesor LAPAYRE:

Por medio del presente, hago constancia de mi participación en la validación médica de una ontología de diagnóstico de las enfermedades vasculares del corazón, cerebro, riñones y del sistema periférico. Mi trabajo de validación consistió en verificar la correcta aplicación de nociones, definiciones, relaciones y propiedades de criterios diagnósticos para los padecimientos vasculares.

Cordialmente,



Dra. Sonia María López Correa
Médico Internista, Maestra en Ciencias de la Salud
Morelia, Michoacán, México.

Centro médico en Ventura Puente #455 Col Chapultepec Norte C.P. 58260, Morelia, Michoacán
Teléfono : (443)3159859

FIGURE 5.11 – Attestation du Docteurs Sonia Maria López Correa pour la validation de COOVADIS



FIGURE 5.12 – Attestation du Docteurs Luis Fernando Sanchez Contreras pour la validation de COOVADIS

CONCLUSION



Dans ce chapitre, nous avons présenté notre expérience de validation de deux de nos ontologies, celle du système vasculaire et celle du diagnostic au regard de trois problématiques étroitement liées : (1) la validation de la pertinence de notre modèle ontologique, (2) la conception et la structuration des connaissances au sein de nos ontologies de spécialité médicale, et (3) l'application de nos ontologies.

Sur le cas d'application de Mr J. (*notre fil rouge*) nous avons montré comment CO-OVADIS a fourni aux trois médecins de Morelia un outil conforme aux ontologies qui permet de tracer le cheminement du diagnostic : c'est la partie validation en milieu clinique de nos travaux qui est indispensable lorsque l'on propose des solutions pour le télé-diagnostic médical.

CONCLUSION DE LA PARTIE II



Dans cette deuxième partie, nous avons présenté nos contributions.

Le chapitre 3 a tout d'abord permis la définition de nos trois ontologies : l'ontologie du système vasculaire, l'ontologie du diagnostic et enfin l'ontologie de la traçabilité. Les deux dernières ontologies sont réellement très novatrices, et permettent la traçabilité.

Le chapitre 4 a été consacré à la définition de notre plateforme COOVADIS : COI-labOrative VAscular DIagnoSis. Développée en mode SaaS (*Software as a Service*) et utilisant les 3 ontologies précédemment définies, elle permet l'aide à la collaboration entre professionnels de santé pour l'interprétation des données l'amélioration et la traçabilité du diagnostic d'un patient. En fin de chapitre, nous avons exposé le cas de Mr J., qui sera notre *fil rouge* pour la suite document, et qui est un patient dont les diagnostics successifs nous ont permis de montrer comment la plateforme web COOVADIS permet de suivre l'évolution d'une pathologie.

Nous avons enfin présenté dans le dernier chapitre de cette partie la validation de notre système. Il s'agit de la validation théorique, et de la validation par des experts à l'aide de tests en milieu clinique. Cette dernière a nécessité la collaboration active de trois médecins mexicains de Morelia impliqués dans les diagnostics dans le domaine cardiovasculaire (1 radiologue, et 2 cardiologues).

CONCLUSION GÉNÉRALE ET PERSPECTIVES

CONCLUSION GÉNÉRALE

Les travaux de cette Thèse s'inscrivent dans le contexte de la télémédecine et plus particulièrement des nouvelles attentes, en matière de législation, des applications de télédiagnostic médical. Nous avons travaillé en amont des problématiques de l'entreprise *Covalia Interactive*, qui est une start-up issue de notre Équipe de Recherche, et qui commercialise le logiciel *CovotemTM* auprès de nombreux hôpitaux français.

Suite à la promulgation de la loi HPST (*Hôpital, Patients, Santé, Territoires*), et aux recommandations de la CNOM (*Commission Nationale de l'Ordre des Médecins*) la traçabilité des applications de télémédecine a été identifiée comme indispensable.

Devant la masse importante des données manipulées dans les dossiers médicaux, la traçabilité n'est pas une tâche facile. Nous avons ainsi choisi d'utiliser les ontologies pour extraire de manière plus efficace les traces des diagnostics médicaux.

La première partie de ce document a permis de présenter deux états de l'art sur les télé-applications collaboratives (en particulier dans le domaine du télédiagnostic médical) d'une part, et d'autre part, sur le domaine des ontologies.

Le premier état de l'art a permis de montrer les évolutions récentes des télé-applications collaboratives, et plus particulièrement les nouvelles exigences dans le domaine de la télémédecine, comme la traçabilité. Les flux de données pris en compte dans un télé-diagnostic étant de taille très importante, il est apparu indispensable de faire appel aux ontologies pour développer de nouveaux outils de traçabilité qui seront capables de trouver de manière efficace et fiable, dans la masse d'informations, les informations pertinentes pour tracer le télé-diagnostic.

Le chapitre suivant de l'état de l'art a permis d'étudier les méthodes de conception, les outils et les langages qui permettent de construire des ontologies. Ces méthodes ont été utilisées pour élaborer les trois ontologies qui constituent le cœur de notre plateforme *COOVADIS COllabOrative VAscular DiagnoSis*.

La deuxième partie du document constitue notre contribution : définition de trois nouvelles ontologies, puis de la plateforme *COOVADIS*.

Nous avons tout d'abord défini trois nouvelles ontologies : l'ontologie du système vasculaire, l'ontologie du diagnostic, et enfin l'ontologie de la traçabilité. Les deux dernières ontologies sont réellement très novatrices, et permettent la traçabilité.

Les chapitres suivants sont consacrés à la présentation de la plateforme *COOVADIS* : architecture de la plateforme développée en mode SAAS *Software As A Service*, et intro-

duction de cas d'utilisation (Mr J. *fil rouge* du document), puis exposition des premiers résultats de traçabilité et évaluation des trois ontologies par les médecins. Nous avons validé les ontologies de manière théorique ainsi qu'avec trois médecins mexicains de Morelia (validation en milieu clinique).

Les résultats sont très encourageants :

- d'un point de vue théorique tout d'abord, nos deux ontologies de traçabilité (ontologie du diagnostic et ontologie de la traçabilité) permettent le traitement direct des diagnostics liés au même patient (ontologie de diagnostic), et elles gèrent la correspondance du travail collaboratif sur des examens concrets, c'est-à-dire, la traçabilité des aides au diagnostic pour les experts en lien avec les données d'examen au format DICOM déjà explorées.
- d'un point de vue applicatif : il faut rappeler que ces travaux s'inscrivent en amont des problématiques R&D de l'entreprise Covalia Interactive. Et dans l'état actuel d'avancement de nos tests, la solution de traçabilité de notre plateforme est très convaincante. En effet, dans une trace, la plateforme COOVADIS fournit toutes les images issues des examens passés par le patient. Et pour chaque image, il est possible d'en afficher les détails et, de façon sémantique : les connaissances qui sont y attachées mais également celles issues de la démarche complète de diagnostic du cas du patient. De plus, la plateforme fournit les informations relatives au patient, au médecin traitant et au type d'imagerie utilisé ainsi qu'à la collaboration pour la prise de décision clinique.

PERSPECTIVES

VALIDATION PAR ESSAIS CLINIQUES

Depuis le début du document, nous parlons de *validation en milieu clinique*. En effet, pour les applications dans le domaine médical, une *validation clinique* est nécessaire mais cela implique des contraintes en particulier de temps qui ne permettait pas de la réaliser au cours de ces travaux. Nous avons commencé sur quelques cas dont celui de Mr J. mais il faudra maintenant passer à la phase réelle de *validation clinique*. Dans un premier temps durant 6 à 12 mois, la plateforme sera utilisée à Morelia, et cette utilisation permettra les dernières mises au point de la plateforme et de ses ontologies. C'est une phase contraignante pour les médecins car elle sera faite en parallèle de la démarche *classique* de diagnostic, elle ne se substituera pas à cette dernière.

Ensuite, il faudra déposer un dossier de demande d'Autorisation d'*Essai Clinique* (AEC) auprès de l'*ANSM* (Agence Nationale de Sécurité du Médicament et des produits de santé) :

- description du type d'investigation,
- protocole d'évaluation de satisfaction de la plateforme,
- nombre de patients,
- source du panel de patients,
- autorisation des hôpitaux où s'effectueront ces essais,
- ...

Et ce n'est qu'à l'issue de toute cette démarche que pourront commencer les *essais cliniques*. Il faudra donc un minimum de patients (habituellement de 100 à 200 patients) ce qui prendra beaucoup de temps (au moins 12 mois).

GÉNÉRICITÉ DE L'ONTOLOGIE DE DIAGNOSTIC

L'ontologie de la pathologie ne peut pas être générique, en effet nous n'échapperons pas au fait de devoir développer une nouvelle ontologie par nouveau domaine médical traité. Concernant l'ontologie de traçabilité, elle est quant à elle indépendante de la pathologie.

Reste l'ontologie de diagnostic qui est liée à la pathologie et qu'il sera intéressant de rendre générique pour faciliter l'adaptabilité de la plateforme, et le traitement de divers pathologies. Pour le développement d'une ontologie de diagnostic générique, il est nécessaire de construire un réseau conceptuel formé par combinaisons des types d'examen propres aux maladies à traiter. Grâce au fait que notre ontologie comprend des terminologies ontologiques de diagnostic vasculaire (types d'imagerie, d'examen physique, de procédures, résultats de laboratoire et mesures, observations complémentaires, coordonnées de patient, coordonnées de médecin traitant, information sur l'institution, ...), les concepts sont annotés avec des propriétés, et modélisés en logique de description au format OWL. Il est possible ainsi d'ajouter des règles associatives et des nouveaux concepts de diagnostic qui permettent d'aborder d'autres pathologies du patient.

Afin de valider un enrichissement d'ontologie dans un domaine diagnostique qui peut comprendre les différents types pathologiques et d'exploration, notre réflexion a porté sur la collecte, l'organisation, la représentation, la formalisation et l'enrichissement des connaissances dans l'ontologie de diagnostic : la recherche de nouveaux concepts, leurs rôles ainsi qu'une phase de placement de ces concepts et relations au sein de l'ontologie originale.

INTÉGRATION AU SEIN DE COVOTEMTM

Très liée à la perspective précédente (sur la généricité de l'ontologie du diagnostic) l'idée finale de nos travaux est l'intégration à la plateforme CovotemTM de l'entreprise Covalia Interactive.

Mais l'entreprise a été certifiée ISO 13485, par l'organisme notifié SGS, sur la conception, le développement et la vente de systèmes de télémédecine. Cette certification spécifique aux organisations des sociétés dans le domaine médical permet d'assurer un niveau élevé de la qualité des produits et garantit une amélioration continue. Cela implique donc de passer par une phase de validation par *essais cliniques* de nouveaux modules de traçabilité qui entreront dans le produit.

Puis, à nouveau, chaque nouvelle ontologie de pathologie devra subir les mêmes phases de validation.

AIDE AU DIAGNOSTIC : INFÉRENCE ET WEB SÉMANTIQUE

Enfin une perspective indépendante du transfert de ces travaux dans les produits de Covalia Interactive.

Il serait intéressant de pouvoir utiliser les ontologies comme aide au diagnostic en utilisant un moteur d'inférences. Il s'agirait par exemple, en utilisant les symptômes issus des phénotypes du patient, d'inférer quelles pourraient être les pathologies possibles. Ces propositions pourraient être formulées aux médecins comme aide à la prise de décision

clinique. Un autre exemple serait de proposer aux médecins un traitement et les effets secondaires lors de la définition du diagnostic.

L'aide au diagnostic pourrait également être améliorée en proposant plus d'informations que celles contenues dans nos ontologies mais recensées dans des réseaux sémantiques comme par exemple UMLS (*Unified Medical Language System*). Ce réseau sémantique pourrait être utilisé en analysant de façon automatique les comptes rendus cardiovasculaires afin de proposer des recommandations en vue d'améliorer la prise de décisions cliniques et la prise en charge des patients. Il faudrait pour cela, indexer nos concepts pour permettre que les réseaux ontologiques travaillent en collaboration.

Enfin, l'aide au diagnostic pourrait être facilité en proposant un système de requêtes aux médecins permettant la recherche de cas similaires à l'aide des symptômes ou des pathologies.

MA BIBLIOGRAPHIE PERSONNELLE

(Y COMPRIS PUBLICATIONS ANTÉRIEURES AU COURS DU MASTER)

ARTICLE EN REVUE

- [San13] Sanchez Santana Maria-Aydee and Aupet Jean-Baptiste and Betbeder Marie-Laure and Lapayre Jean-Christophe and Camarena Antonio. A Tool for Telediagnosis of Cardiovascular Diseases in a Collaborative and Adaptive Approach. JUSC, Journal of Universal Computer Science. ISIWEB IF 0.762, Pages 1275–1294, Vol. 19(9), 2013.

PUBLICATIONS EN CONFÉRENCE AVEC COMITÉ DE LECTURE D'AUDIENCIE INTERNATIONALE, DONT LES ACTES SONT PUBLIÉS

- [San14] Sanchez Santana Maria Aydee and Betbeder Marie-Laure and Lapayre Jean-Christophe and Carranza Madrigal Jaime. COOVADIS : A New Framework for Collaborative Diagnosis Approach of the Vascular System. Proceedings of the CSCWD 2014, 18th IEEE Int. Conf. on Computer Supported Cooperative Work in Design. Pages 627–632, Hsinchu, Taiwan, May 2014.
- [San12] Sanchez Santana Maria Aydee and Aupet Jean-Baptiste and Betbeder Marie-Laure and Lapayre Jean-Christophe and Camarena Ibarrola Jose Antonio. Adaptive Collaborative Environment for Vascular Problems Telediagnosis. Proceedings of the LNCS IWAAL 2012, 4th Int. Workshop on Ambient Assisted Living. Pages 1–8, Vitoria-Gasteiz, Spain, December, 2012.

ARTICLE EN REVUE NATIONAL

- [San12a] Sanchez Santana Maria Aydeé and Carranza Madrigal Jaime. Desarrollo de un programa de cómputo para medir automáticamente el espesor de la intima media carotídea y la luz de la arteria humeral mediante ultrasonografía así como el cálculo del riesgo cardiovascular. Anales de Radiología México Pages 20–26, Vol. 1, 2012. Publiée au cours du Master á Morelia.

PUBLICATIONS EN CONFÉRENCE AVEC COMITÉ DE LECTURE D'AUDIENCIE NATIONALE, DONT LES ACTES SONT PUBLIÉS

[San12b] Sanchez Santana Maria-Aydee and Aupet Jean-Baptiste and Betbeder Marie-Laure and Lapayre Jean-Christophe and Camarena Ibarrola Jose Antonio. Environnement Collaboratif d'Adaptation pour le Télédiagnostic des Problèmes Vasculaires. Ubimob 2012, 8èmes journées francophones Mobilité et Ubiquité. Pages 311–321, Biarritz, France, June 2012.

AUTRES PUBLICATIONS

[San10] Sanchez Santana Maria Aydee and Carranza Madrigal Jaime and Camarena Ibarrola Jose Antonio and Calderon Solorio Felix. Discrepancy Detection of Cardiovascular Abnormalities : Computer System to Manual Measurement by Radiologists. XXXIII National Internal Medicine Congress of México, Vol6(1). Cancún, México. November 2010. Publiée au cours du Master á Morelia.

[San10a] Sanchez Santana Maria Aydee and Carranza Madrigal Jaime. Software for Automatic Diagnosis of Cardiovascular Diseases Through Ultrasonographic Measurement of Arteries. Institute of Electrical and Electronic Engineers IEEE, Morelia Institute of Tecnology. Pages 107–113, Morelia, México. October 2010, Publiée au cours du Master á Morelia.

[San09] Sanchez Santana Maria Aydee and Carranza Madrigal Jaime and Camarena Ibarrola Jose Antonio and Calderon Solorio Felix. Software for Automatic Measurement of Carotid Alterations and Endothelial Function in the Brachial Artery. XXXII National Internal Medicine Congress. Pages 15–19, Vol.25(1), Mérida, México. November 2009. Publiée au cours du Master á Morelia.

[San08] Sanchez Santana Maria Aydee and Carranza Madrigal Jaime. Software for Automatic Measurement of Carotid Plaque and Brachial Diameter. VII State Hypertension Congress. Pages 91–94, Vol.6(2), Morelia, México. May 2008. Publiée au cours du Master á Morelia.

BIBLIOGRAPHIE

- [Aba13] Asma Ben Abacha, Marcos Da Silveira, and Cédric Pruski. *Medical Ontology Validation through Question Answering*, volume 7885, pages 196–205. Springer-Verlag, 2013.
- [Ant09] Erick Antezana, Mikel Egana, Ward Blondé, Aitzol Illarramendi, Inaki Bilbao, Bernard De Baets, Robert Stevens, Vladimir Mironov, and Martin Kuiper. The cell cycle ontology : an application ontology for the representation and integrated analysis of the cell cycle process. *Genome Biology*, 10(5) :1–20, 2009.
- [Arm09] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andrew Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, and Matei Zaharia. Above the clouds : A berkeley view of cloud computing. Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley, February 2009.
- [Ash04] Joan S. Ash, Marc Berg, and Enrico Coiera. Some unintended consequences of information technology in health care : The nature of patient care information system-related errors. *Journal of the American Medical Informatics Association*, 11(2) :104–112, March 2004.
- [Asm12] Pierre Zweigenbaum Asma Ben Abacha. Une étude comparative empirique sur la reconnaissance des entités médicales. In *Traitement Automatique des Langues*, 53(1) :39–68, dec 2012.
- [Asm13] Cédric Pruski Asma Ben Abacha, Marcos Da Silveira. Medical Ontology Validation through Question Answering. In *Proceedings of the 14th conference on AI in Medicine in Europe : Artificial Intelligence Medicine*, AIME'13, pages 196–205, Murcia, Spain, May 2013. Springer-Verlag.
- [Aup10] Jean-Baptiste Aupet, Eric Garcia, Hervé Guyennet, Jean-Christophe Lapayre, and David Martins. Security in medical telediagnosis. In *Book Multimedia Services in Intelligent Environments - Integrated Systems, Chapter 9*, Springer, 2010.
- [Aup12] Jean-Baptiste Aupet. Environnements collaboratifs temps réel : Adaptabilité des échanges de données multimédia. *Thèse de Doctorat*, January 2012.
- [Aus00] Nathalie Aussenac-gilles and Didier Bourigault. The th(ic)2 initiative : Corpus-based thesaurus. In *Workshop on Ontologies and Texts, Juans-Les-Pins (France)*, volume 51 of *EKA'2000*, 2000.
- [Aus05] Nathalie Aussenac-Gilles. *Méthodes ascendantes pour l'ingénierie des connaissances*. PhD thesis, Université Toulouse III, December 2005.
- [Ban05] Audrey Baneyx, Jean Charlet, and Marie-Christine Jaulent. Building medical ontologies based on terminology extraction from texts : an experimentation in pneumology. In *Connecting Medical Informatics and Bio-Informatics*, pages 27–50, Geneva, Switzerland, August 2005. Studies in Health Technology and Informatics.

- [Ban08] Audrey Baneyx and Jean Charlet. Ontopneumo : An ontology of pneumology domain. *Applied Ontology*, 3(4) :229–233, December 2008.
- [Ben02] Khalid Benali, Grégory Bourguin, Bertrand David, Alain Derycke, and Christine Ferraris. Collaboration/coopération. In *Actes des deuxièmes assises nationales du GdR I3*, pages 79–94, 2002.
- [Ber01] Tim Berners-lee and James Hendler. The semantic web. *Scientific American*, 284 :34–43, 2001.
- [Bha09] Mehul Bhatt, Wenny Rahayu, Sury Prakash Soni, and Carlo Wouters. Ontology driven semantic profiling and retrieval in medical information systems. *Web Semant.*, 7(4) :317–331, December 2009.
- [Bie99] Brigitte Biebow and Sylvie Szulman. Terminae : A linguistics-based tool for the building of a domain ontology. In Dieter Fensel and Rudi Studer, editors, *Knowledge Acquisition, Modeling and Management*, volume 1621 of *Lecture Notes in Computer Science*, pages 49–66. Springer Berlin Heidelberg, 1999.
- [Bir12] Henk Birkholz, Ingo Sieverdingbeck, Karsten Sohr, and Carsten Bormann. Io : An interconnected asset ontology in support of risk management processes. In *Proceedings of the 2012 Seventh International Conference on Availability, Reliability and Security*, ARES '12, pages 534–541. IEEE Computer Society, 2012.
- [Bla02] Christian Blaschke and Alfonso Valencia. Automatic ontology construction from the literature. *genome informatics*, 2002.
- [Bla09] Ignacio Blanquer Espert, Vicente Hernández García, Fco. Javier Mesequer Anastásio, and J. Damií Segrelles Quilis. Content-based organisation of virtual repositories of dicom objects. *Future Generation Computer Systems*, 25(6) :627–637, June 2009.
- [Bod06b] Olivier Bodenreider. Lexical, terminological and ontological resources for biological text mining. *Text mining for biology and biomedicine*, 2006.
- [Bod06a] Olivier Bodenreider and Robert Stevens. Bio-ontologies : current trends and future directions. *Briefings in Bioinformatics*, 7(3) :256–274, September 2006.
- [Bre04] Christopher Brewster, Harith Alani, Srinandan Dasmahapatra, and Yorick Wilks. Data driven ontology evaluation. In *International Conference on Language Resources and Evaluation*, pages Lisbon, Portugal, Lisbon, Portugal, May 2004.
- [Cao10] Tru H. Cao and Anh H. Mai. Ontology-based understanding of natural language queries using nested conceptual graphs. In *Proceedings of the 18th International Conference on Conceptual Structures : From Information to Intelligence*, ICCS'10, pages 70–83, Berlin, Heidelberg, 2010. Springer-Verlag.
- [Char02] Jean Charlet. L'ingénierie de connaissance : développements, résultats et perspectives pour la gestion des connaissances médicales, December 2002.
- [Char06] Jean Charlet, Bruno Bachimont, and Marie-Christine Jaulent. Building medical ontologies by terminology extraction from texts : An experiment for the intensive care units. *Comp. in Bio. and Med.*, 36(7-8) :857–870, 2006.
- [Char12] Jean Charlet, Gunnar Declerck, Ferdinand Dhombres, Pierre Gayet, Patrick Miroux, and Pierre-Yves Vandenbussche. Construire une ontologie médicale pour la recherche d'information : problématiques terminologiques et

- de modélisation. In *Journées Francophones d'Ingénierie des Connaissances*, July 2012.
- [Cim00] James J. Cimino. Desiderata for controlled medical vocabularies in the twenty-first century. In *Methods of Information in Medicine*, pages 394–403, 2000.
- [Cor03] Oscar Corcho, Mariano Fernandez-lopez, and Asuncion Gomez-perez. Methodologies, tools and languages for building ontologies. where is their meeting point. *Data & Knowledge Engineering*, 46 :41–64, 2003.
- [Cor05] Oscar Corcho, Mariano Fernández-lópez, Asuncion Gomez-perez, and Angel Lopez. Building legal ontologies with methontology and webode. In *Law and the Semantic Web*, pages 142–157. Springer-Verlag, 2005.
- [Cor06] Ronald Cornet, Nicolette Keizer, and Ameen Abu-Hanna. A framework for characterizing terminological systems. *Methods of Information in Medicine*, 45 :253–266, October 2006.
- [Cru04] Isabel F. Cruz, Huiyong Xiao, and Feihong Hsu. An ontology-based framework for xml semantic integration. In *Proceedings of the International Database Engineering and Applications Symposium, IDEAS '04*, pages 217–226, Washington, DC, USA, 2004. IEEE Computer Society.
- [Dav01] B. David. Ihm pour les collecticiels. In *Revue Réseaux et Systèmes Répartis, Calculateurs Parallèles*, volume 2(3), Ed. Hermès, 2001.
- [Del13] Pari Delir Haghighi, Frada Burstein, Arkady Zaslavsky, and Paul Arbon. Development and evaluation of ontology for intelligent decision support in medical emergency management for mass gatherings. *Decision Support Systems*, 54(2) :1192–1204, January 2013.
- [Dib13] Pierre Dibon, Marc Dalmau, and Philippe Roose. Ubiquitous widgets : Designing interactions architecture for adaptive mobile applications. In *Proceedings of the Int. Conf. Distributed Computing in Sensor Systems (DCOSS)*, pages 331–336, May 2013.
- [Duq09] Rafael Duque, Crescencio Bravo, and Manuel Ortega. A conceptual model for analysing collaborative work and products in groupware systems. In *Proceedings of the 6th international conference on Cooperative design, visualization, and engineering, CDVE'09*, pages 125–132, Berlin, 2009. Springer-Verlag.
- [Ell94] C. Ellis and C. Wainer. A conceptuel model of groupware. In *Proceedings of CSCW'94, ACM Press*, pages 79–88, June 1994.
- [Elm07] Nabil Elmarzouqi, Eric Garcia, and Jean-Christophe Lapayre. Accm : a new architecture model for cscw. In *Procs. of the 10th IEEE and LNCS Int. Conf. on CSCW in Design*, volume 1, pages 84–91, 2007.
- [Fay10] J. Fayn and P. Rubel. Toward a personal health society in cardiology. In *IEEE Trans Inf Technol Biomed.*, volume 14(2), pages 401–409. IEEE Press, March 2010.
- [Fer00] Mariano Fernández López, Asunción Gómez-Pérez, and María Dolores Rojas Amaya. Ontology's crossed life cycles. In *Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management, EKAW '00*, pages 65–79, London, UK, UK, 2000. Springer-Verlag.
- [Fer07] Carlos Ruben Ferreira, Pedro Marques, Andre L. Martins, Sergio Rita, Bruno Grilo, Rudi Araujo, Peyman Sazedj, and H. Sofia Pinto. Ontology design risk

- analysis. In *Proceedings of the 2007 OTM Confederated International Conference on the Move to Meaningful Internet Systems*, OTM'07, pages 522–533, Berlin, Heidelberg, 2007. Springer-Verlag.
- [Fui08] David Fuin, Eric Garcia, Herve Guyennet, and Jean-Christophe Lapayre. Collaborative interactions for medical e-diagnosis. In *HPCN, Int. Journal on High-Performance Computing and Networking*, volume 5(3), pages 189–197, 2008.
- [Gan02] Fabien Gandon. Ontology engineering : a survey and a return on experience, March 2002.
- [Gan06] Fabiel L. Gandon. Ontologies informatiques. In *Security of the Distributed Electronic Patient Record. Conference*, 2006.
- [Gang06] Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, and Jos Lehmann. Modelling ontology evaluation and validation. In *Proceedings of the 3rd European Semantic Web Conference*, ESWC2006, pages 140–154. Springer, 2006.
- [Gar05] Eric Garcia, Herve Guyennet, Jean-Christophe Lapayre, and Thierry Moulin. Adaptive tele-application for remote neurology diagnosis. In *International Journal of Telemedicine and e-Health*, volume 11(6), pages 692–701, 2005.
- [Gas09] Dragan Gasevic, Dragan Djuric, and Vladan Devedzic. *Model Driven Engineering and Ontology Development*. Springer Publishing Company, Incorporated, 2nd edition, 2009.
- [Gei13] Peter Geibel, Martin Trautwein, and al. Ontology-based semantic annotation of documents in the context of patient identification for clinical trials. In *On the Move to Meaningful Internet Systems : OTM 2013 Conferences*, volume 8185 of *Lecture Notes in Computer Science*, pages 719–736. Springer Berlin Heidelberg, 2013.
- [Gol03] Jennifer Golbeck, Gilberto Fragoso, Frank Hartel, Jim Hendler, Jim Oberthaler, and Bijan Parsia. The national cancer institute's thesaurus and ontology. *Web Semantics : Science, Services and Agents on the World Wide Web*, 1(1), 2003.
- [Gom04] Asuncion Gomez-perez, Mariano Fernandez-lopez, and Oscar Corcho. Methodologies and methods for building ontologies. In *Ontological Engineering, Advanced Information and Knowledge Processing*, pages 107–197. Springer London, 2004.
- [Gon11] R. S. Goncalves, B. Parsia, and U. Sattler. Analysing the evolution of the nci thesaurus. In *Proceedings of the 2011 24th International Symposium on Computer-Based Medical Systems*, CBMS '11, pages 1–6, Washington, DC, USA, 2011. IEEE Computer Society.
- [Gom07] asuncion Gomez-Perez, Mariano Fernandez Lopez, and Oscar Corcho. *Ontological Engineering : With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web*. Advanced Information and Knowledge Processing, Secaucus, NJ, USA, springer-verlag edition, 2007.
- [Gru92] Thomas R. Gruber. Ontolingua : A mechanism to support portable ontologies. Technical report, Stanford Knowledge Systems Laboratory, 1992.
- [Gru93b] Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. In *Formal Ontology in Conceptual Analysis and knowledge Representation*, Kluwer Academic Publishers. Kluwer Academic Publishers, 1993.

- [Gru93a] Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2) :199–220, June 1993.
- [Gru95] Michael Grüninger and Mark S. Fox. Methodology for the design and evaluation of ontologies. In *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing*, 1995.
- [Gru09] Tom Gruber. Ontology. In *Encyclopedia of Database Systems*, pages 1963–1965. Springer-Verlag, 2009.
- [Gua98] Nicola Guarino. Formal ontology and information systems. In *Proceedings of the 1st International Conference*, pages 3–15, Trento, Italy, June 1998. IOS Press.
- [Gua09] Nicola Guarino and Christopher A. Welty. An overview of ontoclean. In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*, International Handbooks on Information Systems, pages 201–220. Springer Berlin Heidelberg, 2009.
- [Hak04] A. Hakeem. Ontology and taxonomy collaborated framework for meeting classification. In *Pattern Recognition. Proceedings of the 17th International Conference*, volume 4 of *ICPR 2004*, pages 219–222, 2004.
- [Hor02] Ian Horrocks. Daml+oil : A reason-able web ontology language. In *Advances in Database Technology — EDBT 2002*, volume 2287, pages 2–13. Springer Berlin Heidelberg, 2002.
- [HPST12] loi Hpst. Loi n. 2009-879 du 21 juillet 2009 portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires. In *en ligne Journal officiel*, n. 167 du 22 juillet 2009, page 51, May 2009.
- [Hua08] Zhisheng Huang and Franck Van Harmelen. Using semantic distances for reasoning with inconsistent ontologies. In *Proceedings of the 7th International Semantic Web Conference*, ISWC'08, pages 178–194, Karlsruhe, Germany, October 2008. Springer.
- [Jac06] M. Jacovi, V. Soroka, and al. The chasms of cscw : a citation graph analysis of the cscw conference. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 289–298, ACM Press New York, NY, USA, 2006.
- [Kaz11] Yevgeny Kazakov, Markus Krötzsch, and František Simancík. Concurrent classification of el ontologies. In *Proceedings of the 10th International Conference on The Semantic Web*, ISWC'11, pages 305–320, Berlin, Heidelberg, 2011. Springer-Verlag.
- [Ken00] Robert E. Kent. Conceptual knowledge markup language : An introduction. *Netnomics : Economic Research and Electronic Networking*, 2(2) :139–169, June 2000.
- [Ker06] Fabio Kepler, Christian Paz-trillo, Joselyto Riani, Marcio M. Ribeiro, Leliane Nunes De Barros, and Renata Wassermann. Classifying ontologies. In *Proceedings of the Second Workshop on Ontologies and their Applications (WONTO 2006)*, volume 199, October 2006.
- [Khan13] Muhammad Khurram Khan and Saru Kumari. An authentication scheme for secure access to healthcare services. *Journal of Medical System*, 37(4) :1–12, July 2013.

- [Kif97] Michael Kifer and Georg Lausen. F-logic : a higher-order language for reasoning about objects, inheritance, and scheme. In *Proceedings of the ACM/SIGMOD Int. Symposium on Management of Data*, volume 18(2), pages 134–146. ACM, 1997.
- [Kri12] Reshmy Krishnan, Amir Hussain, and P. C. Sherimon. Retrieval of semantic concepts based on analysis of texts for automatic construction of ontology. In *Proceedings of the 19th International Conference on Neural Information Processing, ICONIP'12*, pages 524–532, Berlin, Heidelberg, 2012. Springer-Verlag.
- [Lee11] Wei-Nchih Lee, Will Bridewell, and Amar K. Das. Comparison of semantic similarity measures for application specific ontology pruning. In *Proceedings of the 2011 IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology, HISB '11*, pages 97–103, Washington, DC, USA, 2011. IEEE Computer Society.
- [LeM02] S. Le Moignon, J. Charlet, D. Bourigault, and Jaulent M.C. Construction d'une ontologie à partir de corpus : expérimentation et validation dans le domaine de la réanimation chirurgicale. In *Actes des 6eme Journées Ingénierie des Connaissances*, pages 229–238, Rouen, France, 2002.
- [LeM04] Pascal Le Mer, Luc Soler, Dominique Pavy, Alain Bernard, Johan Moreau, Didier Mutter, and Jacques Marescaux. Argonaute 3d : A real-time cooperative medical planning software on dsl network. In *MMVR12, 12th Annual Medicine Meets Virtual Reality Conference*, volume 98, pages 203–209, Newport Beach, January 2004.
- [Lep08] Paea Lependu, Dejing Dou, Gwen A. Frishkoff, and Jiawei Rong. Ontology database : A new method for semantic modeling and an application to brain-wave data. In *Proceedings of the 20th International Conference on Scientific and Statistical Database Management, SSDBM '08*, pages 313–330, Berlin, Heidelberg, 2008. Springer-Verlag.
- [Len90] Douglas B. Lenat and R. V. Guha. *Building Large Knowledge-Based Systems : Representation and Inference in the Cyc Project*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1989.
- [Lou09] Christine Louberry, Philippe Roose, and Marc Dalmau. Qos-based design process for pervasive computing applications. In *Proceedings of the 6th International Conference on Mobile Technology Application Systems Mobility*, pages 1–4, New York, 2009. ACM.
- [Luc12] Jacques Lucas. Le secret médical et la esante. In *Numéro spécial secret médical de la revue Médecins*, page 21, nov-dec 2012.
- [Mac89] John McCarthy. Artificial intelligence, logic and formalizing common sense. In *Philosophical Logic and Artificial Intelligence*, pages 161–190. Kluwer Academic, 1989.
- [Mac91] Robert M. MacGregor. Inside the loom description classifier. *SIGART Bull.*, 2(3) :88–92, June 1991.
- [Mar02] Jacques Marescaux, Joel Leroy, Francesco Rubino, Michel Vix, Michele Simone, and Didier Mutter. Transcontinental robot assisted remote telesurgery : Feasibility and potential applications. In *Annal of Surgery*, volume 235, pages 487–492. Ann Surg, April 2002.

- [Mat03] Friedemann Mattern and Peter Sturm. From distributed systems to ubiquitous computing - the state of the art, trends, and prospects of future networked systems. In *In K. Irmischer and K.-P. Fährnrich, editors, Proc. KIVS 2003*, pages 3–25. Springer-Verlag, 2003.
- [Mat10] Ely Edison Matos, Fernanda Campos, Regina Braga, and Daniele Palazzi. Celows : An ontology based framework for the provision of semantic web services related to biological models. *Journal of Biomedical Informatics*, 43(1) :125–136, February 2010.
- [Maz14] Ilaria Mazzanti, Maolo Alessandro, and Antonicelli Roberto. E-health and telemedicine in the elderly : State of the art. *Assistive Technologies : Concepts, Methodologies, Tools, and Applications*, pages 693–704, April 2014.
- [Med12] Juan Miguel Medina, Sergio Jaime-Castillo, and Esther Jiménez. A dicom viewer with flexible image retrieval to support diagnosis and treatment of scoliosis. *Expert Systems with Applications : An International Journal*, 39(10) :8799–8808, August 2012.
- [Min09] Hua Min, Frank J. Manion, Elizabeth Goralczyk, Yu-Ning Wong, Eric Ross, and J. Robert Beck. Integration of prostate cancer clinical data using an ontology. *Journal of Biomedical Informatics*, 42(6) :1035–1045, December 2009.
- [Mot98] Enrico Motta. An overview of the ocml modelling language. In *In Proceedings KEML'98 : 8th Workshop on Knowledge Engineering Methods and Languages*, pages 21–22, 1998.
- [Mot00] Enrico Motta, Simon Buckingham Shum, and John Domingue. Ontology-driven document enrichment : Principles, tools and applications. *International Journal of Human-Computer Studies*, 52(6) :1071–1109, June 2000.
- [Noy00] Natalya Fridman Noy and Mark A. Musen. Prompt : Algorithm and tool for automated ontology merging and alignment. In *AAAI/IAAI*, pages 450–455, 2000.
- [Noy01] Natalya F. Noy and Deborah L. McGuinness. Ontology development 101 : A guide to creating your first ontology. Technical report, Knowledge Systems Laboratory Stanford University, 2001.
- [Obr07] Leo Obrst, Werner Ceusters, Inderjeet Mani, Steve Ray, and Barry Smith. The evaluation of ontologies. In *Semantic Web*, pages 139–158. Springer US, 2007.
- [OJEU13] Official Journal of the European Union. Recommendations on a common framework for a unique device identification system of medical devices in the union. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:099:0017:0024:EN:PDF>, April 2013.
- [Pat09] Jyotishman Pathak, Thomas M. Johnson, and Christopher G. Chute. Survey of modular ontology techniques and their applications in the biomedical domain. *Integr. Comput.-Aided Eng.*, 16(3) :225–242, August 2009.
- [Pie09] Christian Pieralli, Bruno Wacogne, Vincent Bonnans, Philippe Humbert, Pazard Lionel, Franck Marzani, Jean-Christophe Lapayre, and Christophe Lang. Collaborative platform for skin cancer screening and associated optical fibered probe for diagnosis. In *Sin'Fran09, Singaporean-French IPAL Symposium*, pages 44–55. World Scientific, February 2009.

- [Pit13] Perrine Pittet, Christophe Cruz, and Christophe Nicolle. Modeling changes for shoin(d) ontologies : An exhaustive structural model. In *Proceedings of the 2013 IEEE Seventh International Conference on Semantic Computing, ICSC '13*, pages 104–109, Washington, DC, USA, 2013. IEEE Computer Society.
- [Por04] Robert Porzel and Rainer Malaka. A task-based approach for ontoly evaluation. In *ECAI Workshop on Ontology Learning and Population*, Valencia, Spain, 2004.
- [Pou12] M. Poulymenopoulou, F. Malamateniou, and G. Vassilacopoulos. Emergency healthcare process automation using mobile computing and cloud services. *Journal of Medical Systems*, 36(5) :3233–3241, October 2012.
- [Rec06] Al. Rector, R. Qamar, and T. Marley. Binding ontologies and coding systems to electronic health records and messages. In Bodenreider O., editor, *Proceedings of the Second International Workshop on Formal Biomedical Knowledge Representation*, pages 11–19, 2006.
- [Rou10] Mohamed Rouane-Hacene, Schahrazed Fennouh, Roger Nkambou, and Petko Valtchev. Refactoring of ontologies : Improving the design of ontological models with concept analysis. In *Proceedings of the 2010 22Nd IEEE International Conference on Tools with Artificial Intelligence*, volume 2 of *ICTAI '10*, pages 167–172, Washington, DC, USA, 2010. IEEE Computer Society.
- [Rui11] Juana Maria Ruiz-Martínez, Rafael Valencia-García, Jesualdo Tomas Fernandez-Breis, Francisco Garcia-Sanchez, and Rodrigo Martinez-Bejar. Ontology learning from biomedical natural language documents using umls. *Expert Systems with Applications*, 38(10) :12365–12378, 2011.
- [Rum05] P. Rumeau, M. Savoldelli, C. Magne, JL. Weber, A. Alonso, F. Castanié, and L. Lareng. Ur-safe project : improving response to telealarm users needs. In *Poster in International Association of Gerontology Congress*, 2005.
- [Sad12] Kazem Sadegh-Zadeh. Medical ontology. In *Handbook of Analytic Philosophy of Medicine*, Philosophy and Medicine, pages 711–756. Springer Netherlands, 2012.
- [Saf00] Charles Safran and Goldberg. Howard. Electronic patient records and the impact of the internet. *International journal of medical informatics*, 60(2) :77–83, June 2000.
- [Sal95] D. Decouchant D. Salber, J. Coutaz and M. Riveill. De l'observabilité et de l'honnêteté : le cas du contrôle d'accès dans la communication homme-homme médiatisée. *Journées IHM'95, Cépaduès*, 1995.
- [Sal12] José Salavert Torres, Damian J. Segrelles Quilis, Ignacio Blanquer Espert, and Vicente Hernandez García. Improving knowledge management through the support of image examination and data annotation using dicom structured reporting. *Journal of Biomedical Informatics*, 45(6) :1066–1074, December 2012.
- [San12] David Sánchez, Albert Sole-Ribalta, Montserrat Batet, and Francesc Serratosa. Enabling semantic similarity estimation across multiple ontologies : An evaluation in the biomedical domain. *Journal of Biomedical Informatics*, 45(1) :141–155, February 2012.
- [Sar11] K. Saruladha, G. Aghila, and A. Bhuvaneswary. Information content based semantic similarity approaches for multiple biomedical ontologies. In Ajith

- Abraham, Jaime Lloret Mauri, JohnF. Buford, Junichi Suzuki, and Sabu M. Thampi, editors, *Advances in Computing and Communications*, volume 191 of *Communications in Computer and Information Science*, pages 327–336. Springer Berlin Heidelberg, 2011.
- [Sch07] Stefan Schulz and Ingvar Johansson. Continua in biological systems. *The Monist*, pages 499–515, 2007.
- [Sch12] Andrea C. Schalley. Ontology and the lexicon : A natural language processing perspective. *Language Resources and Evaluation*, 46(1) :95–100, March 2012.
- [Smi04] Barry Smith. Beyond concepts : Ontology as reality representation. In A. C. Varzi and L. Vieu, editors, *Formal Ontology in Information Systems*, pages 73–84. IOS Press, 2004.
- [Sta09] Steffen Staab and Rudi Studer. *Handbook on Ontologies*. Springer Publishing Company, Incorporated, 2nd ed. edition, 2009.
- [Ste07] O. Steichen, C. Daniel-Le Bozec, M.C. Jaulent, and J. Charlet. Construction d'une ontologie pour la prise en charge de l'hypertension artérielle. *Actes des 18ème Journées en Ingénierie des Connaissances*, pages 241–253, June 2007.
- [Ste11] Michael Kleiber and Thomas Alexander. Evaluation of a mobile ar tele-maintenance system. In *Universal Access in Human-Computer Interaction. Applications and Services*, volume 6768 of *Lecture Notes in Computer Science*, pages 253–262. Springer, 2011.
- [Stu98] R. Studer, V.R. Benjamins, and D. Fensel. Knowledge engineering : Principles and methods. *IEEE Transactions on Data and Knowledge Engineering*, 25(162) :161–197, 1998.
- [Sua12] MariCarmen Suárez-Figueroa, Asunción Gómez-Pérez, and Mariano Fernández-López. The neon methodology for ontology engineering. In *Ontology Engineering in a Networked World*, pages 9–34. Springer Berlin Heidelberg, 12.
- [Sul14] Nabil Sultan. Making use of cloud computing for healthcare provision : Opportunities and challenges. *International Journal of Information Management : The Journal for Information*, 34(2) :177–184, April 2014.
- [Tar10] Samir Tartir, I. Budak Arpinar, and Amit P. Sheth. Ontological evaluation and validation. In *Theory and Applications of Ontology : Computer Applications*, pages 115–130. Springer Netherlands, 2010.
- [Tho11] Janke Thomas. Ontology engineering based on domain specific languages and the application of ontology design patterns. In *Advanced Information Systems Engineering Workshops*, volume 83 of *Lecture Notes in Business Information Processing*, pages 167–176. Springer Berlin Heidelberg, 2011.
- [Uli04] Mihaela Ulieru. *Internet-Enabled Soft Computing Holarchies for e-Health Applications*, pages 131–165. Enhancing the Power of the Internet. Springer-Verlag, 2004.
- [Ush95] Mike Uschold and Martin King. Towards a methodology for building ontologies. In *In Workshop on Basic Ontological Issues in Knowledge Sharing*, EATIS '07, 1995.

- [Val05] Rafael Valencia. *Un Entorno para la Extraccion Incremental de Conocimiento desde Texto en Lenguaje Natural*. PhD thesis, Departamento de la Informática y las Telecomunicaciones, Universidad de Murcia, Murcia, Spain, April 2005.
- [Van11] Klaar Vanopstal, Robert Stichele, Godelieve Laureys, and Joost Buysschaert. Vocabularies and retrieval tools in biomedicine : Disentangling the terminological knot. *J. Med. Syst.*, 35(4) :527–543, August 2011.
- [Vii04] Mika Viinikkala. *Ontology in information systems*, March 2004.
- [Vor10] Tim Vor der Bruck and Holger Stenzhorn. Logical ontology validation using an automatic theorem prover. In *Proceedings of the 2010 Conference on ECAI 2010 : 19th European Conference on Artificial Intelligence*, pages 491–496, Amsterdam, The Netherlands, 2010. IOS Press.
- [Vos07] Michal Vossberg, Dagmar Krefting, and Thomas Tolxdorff. Using dicom in medical grids : Secure image communication and integration of external dicom devices in globus grids. In *Proceedings of the Fifth IASTED International Conference : Biomedical Engineering*, BIEN '07, pages 320–325, Anaheim, CA, USA, 2007. ACTA Press.
- [W3C13] Semantic web (w3c). <http://www.w3.org/standards/semanticweb/>, 2013.
- [Wat10] Ryan Watkins. e-learning - tool for training and professional development services, e-learning, development of knowledge and/or skills for building competence. In *Handbook of Improving Performance in the Workplace : Selecting and Implementing Performance Interventions*, pages 577–597. John Wiley and Sons, Inc., 2010.
- [Weic09] Albert Weichselbraun, Gerhard Wohlgenannt, Arno Scharl, Michael Granitzer, Thomas Neidhart, and Andreas Juffinger. Discovery and evaluation of non-taxonomic relations in domain ontologies. *International Journal Metadata Semantic Ontologies*, 4(3) :212–222, August 2009.
- [Wei09] Michael Weiss. Resource description framework. In *Encyclopedia of Database Systems*, pages 2423–2425. Springer US, 2009.
- [Yun11] Hongyan Yun, Jianliang Xu, Jing Xiong, and Moji Wei. A knowledge engineering approach to develop domain ontology. *International Journal of Distance Education Technologies*, 9(1) :57–71, January 2011.
- [Zom10] Zsolt Zombori. Two phase description logic reasoning for efficient information retrieval. In *Proceedings of the 7th International Conference on The Semantic Web : Research and Applications*, ESWC'10, pages 498–502, Berlin, Heidelberg, 2010. Springer-Verlag.
- [Zwe99] Pierre Zweigenbaum. Encoder l'information médicale : des terminologies aux systèmes de représentation des connaissances. In *Innovation Stratégique en Information de Santé (ISIS)*, pages 27–47, 1999.

TABLE DES FIGURES

1	Schéma global d'organisation des travaux de recherche	16
1.1	Place du collectif dans le domaine du CSCW	24
1.2	Trèfle fonctionnel enrichi	26
1.3	Cloud Computing	28
1.4	Covotem : une plateforme de télédiagnostic temps réel sécurisée	30
1.5	Da Vinci : Robot chirurgical avec distanciation	31
1.6	Holarchie médicale	32
2.1	Classification des ontologies pour le type de problème à résoudre	38
2.2	Processus de développement d'ontologie de Methontology	41
2.3	Processus de développement d'ontologie de NeOn [Sua12]	42
2.4	L'architecture de niveaux de Berners Lee	43
2.5	Architecture de connaissance pour la construction d'une ontologie médicale	47
2.6	Connaissances épistémologiques de l'ontologie	53
2.7	Services de terminologie en médiation entre l'ontologie et l'utilisateur final .	54
2.8	Les classes de maladies/traitement dans une ontologie	56
2.9	Les classes de maladies dans une ontologie	56
2.10	Les classes de traitement dans une ontologie	57
2.11	Les classes de condition de maladie dans une ontologie	57
2.12	Classification des ontologies selon Guarino [Gua98]	58
2.13	Ontologies existantes dans le domaine médical	59
2.14	Cycle de vie d'une ontologie	62
3.1	Processus de développement d'athérosclérose par dysfonction endothéliale	74
3.2	Description de la méthode de développement de l'ontologie	76
3.3	Tâches de conceptualisation d'ontologie	78
3.4	Étapes générales de développement d'une ontologique	80
3.5	Graphe conceptuel général du système vasculaire	85
3.6	Classe OWL du concept « <i>heart valve disease</i> » (pathologies du cœur) . .	91
3.7	Classe OWL du concept Infarctus comme pathologies du cœur et du cerveau	92

3.8	Modèle spécifique du concept Infarctus	93
3.9	Implémentation dans protégé de l'ontologie du système vasculaire	94
3.10	Ontologie du système vasculaire	95
3.11	Structure générale d'un système d'archivage PACS	96
3.12	Génération d'une étude sur un modèle de DICOM	98
3.13	Conception des données et flux ontologique	99
3.14	Ontologie conceptuelle du diagnostic	99
3.15	Liaisons entre maladies vasculaires et diagnostic	100
3.16	Implémentation dans protégé de l'ontologie du diagnostic	102
3.17	Ontologie du diagnostic	103
3.18	Le travail collaboratif sur l'amélioration du diagnostic	104
3.19	Diagramme conceptuel de l'ontologie de traçabilité	106
3.20	Diagramme conceptuel de l'ontologie de traçabilité	107
3.21	Composants de l'ontologie de traçabilité	107
3.22	Implémentation dans protégé de l'ontologie de traçabilité	108
3.23	Ontologie de traçabilité	108
4.1	Architecture globale de COOVADIS	112
4.2	Partage de l'information entre professionnels pour l'amélioration des diagnostics	114
4.3	Connexion des ontologies dans le plateforme COOVADIS	115
4.4	Travail collaboratif pour l'ajout de connaissance sur des cas cliniques	116
4.5	Diagnostics numérisés sur les standards DICOM	117
4.6	Interactions dans la plateforme COOVADIS	119
4.7	Méta-données d'une image DICOM	120
4.8	Union des ontologies	122
4.9	Exemple d'association des concepts	123
4.10	Représentation des connaissances intégrées par les ontologies	124
4.11	La page principale du portail web	125
4.12	La fenêtre d'affichage des connaissances liées aux diagnostics	126
4.13	La fenêtre de collaboration au diagnostic	127
4.14	Saisie du dossier de Mr J.	129
4.15	Méta-données des dossiers médicaux en image Dicom	130
4.16	Concepts et relations	131
5.1	Le processus cyclique de vie dans une ontologie	134

5.2	Validation de la conceptualisation ontologique pour une approche de questions-réponses interactives	137
5.3	Processus de validation	140
5.4	Changement basique	141
5.5	Changement complexe	142
5.6	Changement par suppression	143
5.7	Cas de Mr J. : Résultats de Diagnostic sur COOVADIS	144
5.8	Cas de Mr J. : Résultats avec enrichissement de métadonnées de diagnostic	145
5.9	Lettre de collaboration avec le Dr. Jaime Carranza Madrigal	147
5.10	Attestation du Dr. Jaime Carranza Madrigal pour la validation de COOVADIS	148
5.11	Attestation du Docteurs Sonia Maria López Correa pour la validation de COOVADIS	149
5.12	Attestation du Docteurs Luis Fernando Sanchez Contreras pour la validation de COOVADIS	150

LISTE DES TABLES

3.1 Règles de syntaxe de base pour LD	86
---	----

Résumé :

En médecine, le diagnostic est la démarche par laquelle le médecin, généraliste ou spécialiste va déterminer l'affection dont souffre le patient, et qui va permettre de proposer un traitement. Il repose sur la recherche des causes (pathologie) et des effets (symptômes) de l'affection. Un diagnostic médical efficace doit aujourd'hui intégrer des analyses multidisciplinaires tant au niveau des données que des experts : et compte tenu de la répartition géographique (par exemple de la désertification médicale), il peut être compliqué de réunir au même endroit les experts.

L'évolution des technologies de communication, en particulier Internet, a ouvert de nouvelles possibilités dans le domaine des applications collaboratives à distance et tout particulièrement celui du télé-diagnostic médical : par exemple un panel d'experts distants se réunit virtuellement par l'intermédiaire d'une salle d'examen virtuelle qui favorisera la collaboration afin de coproduire un diagnostic. Mais dans le domaine de la médecine, l'aspect médico-légal est crucial, et il a freiné le développement de ces pratiques à distances.

Dans ce contexte, nous avons développé une plateforme appelée COOVADIS (COllabOrative VAScular DiagnoSis) qui permet la traçabilité dans de telles applications en s'appuyant sur trois ontologies originales (ontologie de la pathologie, ontologie du diagnostic et ontologie de traçabilité). Cette plateforme d'aide à la collaboration entre professionnels de santé a été implémentée en mode SaaS (Software as a Service) sous la forme d'un serveur Web, et validé d'un point de vue théorique et clinique.

Mots-clés : Diagnostic Médical, Télé-médecine, Ontologies, Travail Collaboratif, Traçabilité

Abstract:

In medicine, physicians (general practitioner or specialist) realize a diagnosis to determine patients' disease and propose an adapted treatment. This diagnosis is based on research of causes (pathologies) and effects (symptoms) of affection. Today, to realize an effective medical diagnosis, it is important to realize a multidisciplinary analysis at a data level. But it is also important to make work together experts from different domains. A problem can happen if these experts do not work in the same place. Thus, how is it possible to ease the way to collaborate together?

With evolutions of communication technologies and more particularly Internet, it is easier to develop remote collaborative applications. One of the fields covered by these applications is telemedicine and tediagnosis. Thus, a remote panel of experts can meet together virtually through a virtual room to ease diagnosis collaboration and co-production. Despite everything, forensic aspects slowed down development of remote practices due to privacy and personal information sharing.

In this context, we developed a platform called COOVADIS (COllabOrative VAScular DiagnoSis) that enables traceability in such applications based on three original ontologies (pathologies ontology, diagnosis ontology and traceability ontology). This framework was implemented in SaaS (Software as a Service) as a web server, to support the collaborative work between health professionals. It was also validated from a theoretical and clinical point of view.

Keywords: Medial Diagnosis, Telemedicine, Ontology, Collaborative Work, Traceability

SPIM